Martin Gardner on Math Games!

SPACE DEBRIS • DNA COMPUTERS • SEAWATER IRRIGATION

SCIENTIFIC AMERICAN

AUGUST 1998 $4.95 U.K. £3.20

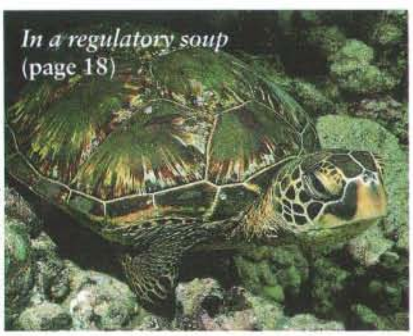
THE Z MACHINE:

PINCHING PLASMA FOR FUSION ENERGY

New Thinking about Back Pain

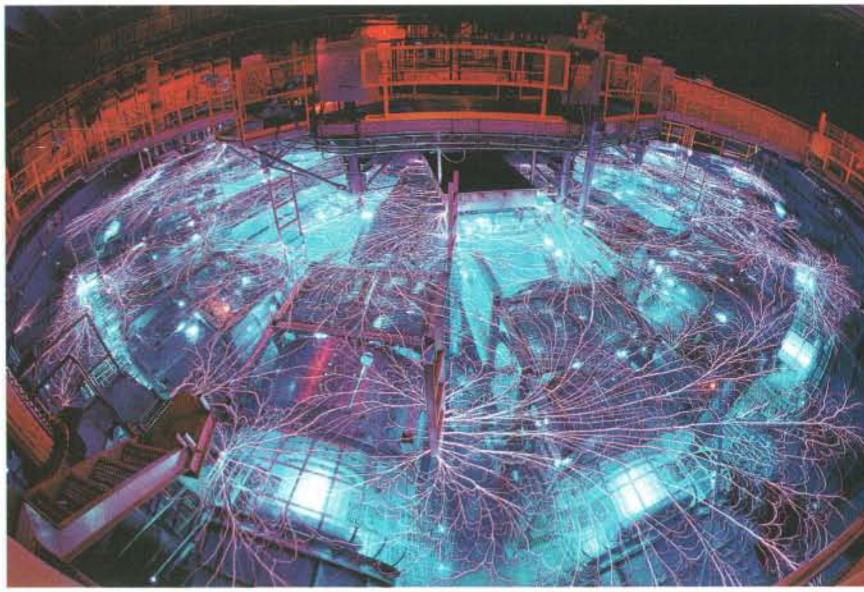
08>
9 770036 873046

SCIENTIFIC AMERICAN
August 1998 Volume 279 Number 2



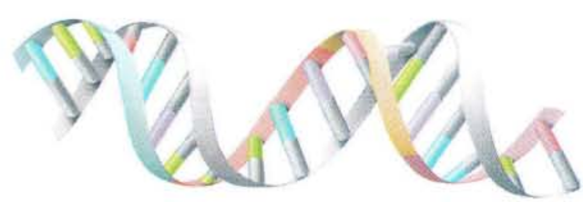
In a regulatory soup (page 18)

CARL MILLER Peter Arnold, Inc.



Even proponents of nuclear fusion research have grown weary of perennial predictions, going back for decades, that fusion power is just 10 years away. But the Z machine, a new device for generating intense pulses of x-rays at Sandia National Laboratories, might finally give that claim credibility again.

Up to 80 percent of all adults eventually suffer back pain. Its possible causes are multifarious and mysterious: why some people experience it is as hard to understand as why many others don't. Fortunately, treatment options are improving, and they usually involve neither surgery nor bed rest.



Electrons moving through silicon do not have a monopoly on computing. Snippets of DNA in a test tube, by combining, growing and recombining, can also solve computational problems. The inventor of DNA computing reminisces about his discovery and discusses its future.

Scientific American (ISSN 0036-8733), published monthly by Scientific American, Inc., 415 Madison Avenue, New York, N.Y. 10017-1111. Periodicals postage paid at New York, N.Y., and at additional mailing offices. Canada Post International Publications Mail (Canadian Distribution) Sales Agreement No. 242764. Canadian BN No. 127387652RT; QST No. Q1015332537. Subscription rates: one year $34.97 (outside U.S. $49). Institutional price: one year $39.95 (outside U.S. $50.95). Postmaster: Send address changes to Scientific American, Box 3187, Harlan, Iowa 51537. Reprints available: write Reprint Department, Scientific American, Inc., 415 Madison Avenue, New York, N.Y. 10017-1111; fax: (212) 355-0408 or send e-mail to sacust@sciam.com **Subscription inquiries: U.S. and Canada (800) 333-1199; other (515) 247-7631.**

About the Cover
The causes of low-back pain remain a great puzzle: although rates of manual labor have fallen, the incidence of backache has risen. *The Thinker* statue by Auguste Rodin; photograph courtesy of the Bridgeman Art Library. Digital manipulation by Jana Brenning.

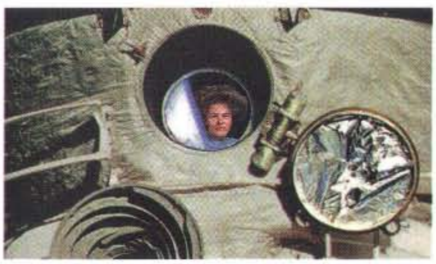
Visit the Scientific American Web site (http://www.sciam.com) for more information on articles and other on-line features.

FROM THE EDITORS

Believing Your Eyes

This is the sort of thing I'd expect of the supermarket tabloids, not of SCIENTIFIC AMERICAN," one letter writer growled. The target of his ire was the cover of our May issue, which portrays astronaut Shannon W. Lucid gazing out a porthole of the Mir space station. He was annoyed at the realism of the picture and at what he mistook to be our intention to fool readers that it was a photograph.

Sorry, but innocent as charged. The "About the Cover" description on that issue's table of contents states that the image is an artist's conception—a fact nearly all readers, I believe, understood. But I understand that writer's confusion, and it gives me the opportunity to discuss the covenant that SCIENTIFIC AMERICAN keeps with its readers.



DIGITAL IMAGES of Shannon Lucid on board Mir (left) *and a Martian sunset* (below) *are artists' interpretations. That fact was noted in the May and July issues, respectively. Drawing on real photographs, these images evoke scenes that no one has witnessed.*

Computers gave artists a new canvas and palette. With scans of photographs, modeling software and an array of digital techniques, a skilled artist can create amazingly lifelike images. For the illustrative purposes of a magazine like this one, digital artistry can be a godsend. It can reimagine unwitnessed scenes (as when Lucid looked out of Mir, or our sunset panorama of Mars from the July issue, both reproduced above); it can display inventions still on the drawing board; it can show individual molecules and other objects that evade regular photography. Conventional art techniques can do the same, but sometimes the digital ones do the job more accurately and realistically. The only catch is that all the realism can mislead readers into accepting fabulous images too literally.

These pros and cons vex editors and art directors at all publications these days, and they all wrestle with the problem in their own way. SCIENTIFIC AMERICAN's policy is and has been to try scrupulously to avoid misleading anyone. When photolike images are computerized creations, we note that fact in the captions and credits appearing alongside. We value our readers' trust too highly to abuse it.

JOHN RENNIE, *Editor in Chief*
editors@sciam.com

DON DIXON (top); U.S. GEOLOGICAL SURVEY AND NASA/JET PROPULSION LABORATORY; SUNSET SIMULATION BY LAURIE GRACE (bottom)

SCIENTIFIC AMERICAN®
Established 1845

SCIENTIFIC AMERICAN, INC.
415 Madison Avenue
New York, NY 10017-1111
(212) 754-0550

PRINTED IN U.S.A.

LETTERS TO THE EDITORS

Readers of the April issue were certainly in the April Fools' spirit. Our deliberately funny tribute to Rube Goldberg received considerable mail, as did John Rennie's editorial on how SCIENTIFIC AMERICAN *really works. Other readers found humor in unexpected places, including a photograph of what seemed to be a guy with lightbulbs in his pants (page 15), and in the article about how females choose their mates.*

TRADE SECRETS

I have to admit that the April Fools' page, "Self-Operating Napkin," by Rube Goldberg [Working Knowledge, April], completely took me in. I must have misread the heading as "self operating system." I believed that the illustration really did divulge part of the workings of the Windows 95 operating system. On second thought, are you sure it doesn't?

A. PETER BLICHER
Princeton, N.J.

BRAVEHEART

"How Females Choose Their Mates," by Lee Alan Dugatkin and Jean-Guy J. Godin [April], reports that female guppies prefer to mate with the male expressing the greatest courage by swimming closest to a predator. The selection advantage of this rash behavior in human male survivors was noted in *Iolanthe,* by Sir William S. Gilbert, in his couplets:

> In for a penny, in for a pound—
> It's Love that makes the
> world go round....
> Beard the lion in his lair—
> None but the brave deserve
> the fair!

DAVID MAGE
Research Triangle Park, N.C.

ACROSS THE SPECTRUM

As a cellular engineer, I enjoyed the special section in the April issue on "Wireless Technologies." In the article "Spread-Spectrum Radio," by David R. Hughes and Dewayne Hendricks, the introduction implies that advances in spread-spectrum radio obviate the need for spectrum conservation and a separate cellular service provider (the "middleman"). But to send signals to one of today's cellular phones, for instance, you need to know its code. The codes are fairly long, so the memory required to store the codes for a large metropolitan area is prohibitive for a single phone. The solution is to store these codes in a central server to which the cellular phone has access. The "middleman" is still a necessity.

TONY DEAN
Motorola, Inc.


SLIM FILMS

COMMUNICATIONS SATELLITE is part of today's wireless network.

Hughes and Hendricks reply:

Dean's comments are quite true for the devices used in today's cellular telephone networks. What we were referring to in our introduction was the technology used in today's Internet, which allows routing information to reside at many different points in a network (rather than on one centralized server), wherever it is necessary and appropriate.

Another good example of this approach would be the use of TCP/IP protocols in amateur radio (a.k.a. ham radio) packet networks, which allow a metropolitan-area network to be deployed and operated without the need of a centralized server, or "middleman."

THE HIGH-TECH LIFE

Marguerite Holloway's profile of Sherry Turkle ["An Ethnologist in Cyberspace," News and Analysis, April] contains the following astonishing sentence: "Turkle says her own childhood was relatively technology free." As she was growing up in Brooklyn in the 1950s and '60s, surely Turkle had electricity, telephones, radio, television, paved streets, automobiles, subway trains and skyscrapers. She probably even crossed some of New York's magnificent bridges on occasion. Today there is a tendency to think of "technology" as anything that is built around an integrated circuit on a chip, like a computer, pager or cell phone, and to take all the rest for granted as though no technology were involved.

S. J. DEITCHMAN
Chevy Chase, Md.

POLIO AFTERMATH

As I was reading Lauro S. Halstead's "Post-Polio Syndrome" [April], his reminder that perhaps only one in 50 of those originally infected with the virus suffered contemporaneous paralysis led me to wonder whether the other 49 might also later be candidates for some variety of post-polio syndrome, albeit not recognized by that name. For instance, do sufferers of chronic fatigue syndrome have their spinal fluid checked for RNA fragments reminiscent of the poliovirus? There is a large group of people who suffered polio, but an even larger group has since received the attenuated virus in the Sabin oral vaccine. Could some ailments in these people be related to post-polio syndrome?

C. G. MADSEN
via e-mail

Halstead replies:

There has been considerable interest over the years in the possible role of poliovirus associated with other diseases. Chronic fatigue syndrome (CFS) is the latest. Although the cause of CFS is still unknown, some researchers believe a virus may be involved, in particular the enteroviruses, which include poliovirus. Unfortunately, most studies of CFS do

not include analyses of patients' spinal fluid. One recent investigation did examine the cerebrospinal fluid, along with other bodily tissues and fluids, and found no evidence of a persistent enterovirus infection.

We continue to see a number of people every year in the clinic with no history of paralysis or polio who develop classic post-polio syndrome. It may turn out that post-polio syndrome is the most common unexpected condition experienced by the 49 individuals who did not suffer paralysis—but unknowingly did sustain a critical threshold of spinal cord damage.

OOPS!

There was an error in the Errata in May: you wrote that the "correct number of platelets in human blood is between 150,000 and 400,000 per cubic centimeter of blood." The correct unit is cubic *millimeter* of blood.

PHRIXOS OVIDIU XENAKIS
San Antonio, Tex.

Editors' note:

We certainly goofed on this one. To set the record straight: in the article "The Search for Blood Substitutes" [February], the chart on page 62 does not contain an error. The article, however, does. On page 61, center column, the text should refer to a cubic millimeter of blood, not a cubic centimeter. We apologize for the errors.

Letters to the editors should be sent by e-mail to editors@sciam.com or by post to Scientific American, 415 Madison Ave., New York, NY 10017. Letters may be edited for length and clarity. Because of the considerable volume of mail received, we cannot answer all correspondence.

ERRATA

In "Nanolasers" [March], one of the researchers who proposed "thresholdless" lasers was misidentified: he is Tetsuro Kobayashi. In "Laser Scissors and Tweezers" [April], the two photographs labeled mouse eggs on page 54 actually show bovine eggs and were taken by Cell Robotics in Albuquerque, N.M.

OTHER EDITIONS OF SCIENTIFIC AMERICAN

Le Scienze
Piazza della Repubblica, 8
20121 Milano, ITALY
tel: +39-2-655-4335
redazione@lescienze.it

Spektrum der Wissenschaft
Verlagsgesellschaft mbH
Vangerowstrasse 20
69115 Heidelberg, GERMANY
tel: +49-6221-50460
redaktion@spektrum.com

Investigacion y Ciencia
Prensa Científica, S.A.
Muntaner, 339 pral. 1.ª
08021 Barcelona, SPAIN
tel: +34-3-4143344
precisa@abaforum.es

Pour la Science
Éditions Belin
8, rue Férou
75006 Paris, FRANCE
tel: +33-1-55-42-84-00

Majallat Al-Oloom
Kuwait Foundation for the Advancement of Sciences
P.O. Box 20856
Safat 13069, KUWAIT
tel: +965-2428186

Swiat Nauki
Proszynski i Ska S.A.
ul. Garazowa 7
02-651 Warszawa, POLAND
tel: +48-022-607-76-40
swiatnauki@proszynski.com.pl

Nikkei Science, Inc.
1-9-5 Otemachi, Chiyoda-Ku
Tokyo 100-66, JAPAN
tel: +813-5255-2800

Svit Nauky
Lviv State Medical University
69 Pekarska Street
290010, Lviv, UKRAINE
tel: +380-322-755856
zavadka@meduniv.lviv.ua

Ke Xue
Institute of Scientific and Technical Information of China
P.O. Box 2104
Chongqing, Sichuan
PEOPLE'S REPUBLIC OF CHINA
tel: +86-236-3863170

50, 100 AND 150 YEARS AGO

AUGUST 1948

SPELLING BEE—"Karl von Frisch, in his studies of honey bees, discovered that bees returning from a rich source of food perform special movements for other bees, which he called dancing. He distinguished between two types of dance: the circling dance (*Rundtanz*) where the bee circles, and the wagging dance (*Schwänzeltanz*) when it moves forward, wagging its abdomen, and turns. Von Frisch showed that the circling dance is used when the food source is closer than about 100 meters. In the wagging dance the frequency of turns indicates the distance to the food source. When the feeding place was 100 meters away, the bee made about 10 short turns in 15 seconds. To indicate a distance of 3,000 meters, it made only three long ones in the same time."

HYPERTENSION—"There is a growing concern over vulnerability to high blood pressure and hardening of the arteries. Death certificates show that these associated conditions kill some 600,000 people in this country annually. Since the 18th century, life expectancy in the U.S. has increased from 39 to 57 years, largely because of the conquest of diseases such as smallpox, typhoid fever, tuberculosis, plague, diphtheria and, more recently, pneumonia and streptococcic infections. The reduction of these diseases has permitted people to live to the age when hypertension and arteriosclerosis take their greatest toll."

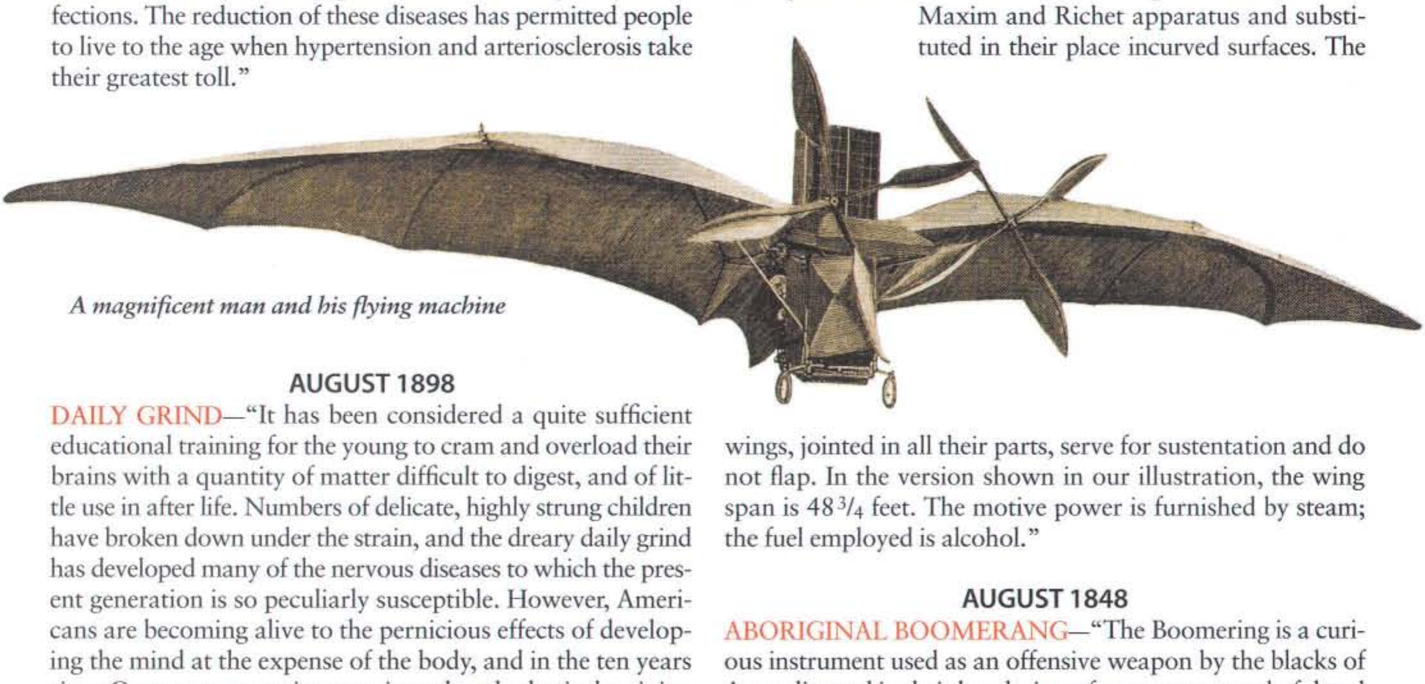
AUGUST 1898

DAILY GRIND—"It has been considered a quite sufficient educational training for the young to cram and overload their brains with a quantity of matter difficult to digest, and of little use in after life. Numbers of delicate, highly strung children have broken down under the strain, and the dreary daily grind has developed many of the nervous diseases to which the present generation is so peculiarly susceptible. However, Americans are becoming alive to the pernicious effects of developing the mind at the expense of the body, and in the ten years since German gymnastics were introduced, physical training holds a place in the curriculum of most larger schools."

EARLIEST MULTIPLEXING?—"Experiments are being conducted on the Paris-Bordeaux line by the inventor M. Mercadier. With some interesting instruments, called duodecaplex, twelve Morse transmitters can work simultaneously on a single wire, each sending its signals to the proper receiver at the end of the line. Each transmitter receives a current through a tuning fork having a special note, its vibrations being electrically maintained. These vibrations cause resonance at the proper receiving circuit. The sifting out of signals, it seems, is very perfect, each receiver giving no evidence of those signals not intended for it." [*Editors' note: This device presages the multiplexing technologies used in analog and digital voice systems.*]

BETTER OPTICS—"The new Zeiss binocular field glasses, now being manufactured in this country by the Bausch & Lomb Optical Company, Rochester, N.Y., are the invention of Prof. Ernst Abbe, of Jena, to whom optical science owes so many recent improvements. The three principal defects of the ordinary field glass are overcome by the use of two pairs of prisms, which erect the inverted image formed by the object glass, shorten the telescope by two-thirds and place the object glasses farther apart than the eyepieces are, thus increasing the stereoscopic effect."

A BIRD-PLANE IDEA—"Is the 'Avion,' an apparatus devised and constructed by M. Clément Ader, a French engineer, finally to permit man to realize the legendary dream of Icarus? Perhaps so. M. Ader has abandoned the plane surfaces of the Maxim and Richet apparatus and substituted in their place incurved surfaces. The wings, jointed in all their parts, serve for sustentation and do not flap. In the version shown in our illustration, the wing span is $48\,^{3}/_{4}$ feet. The motive power is furnished by steam; the fuel employed is alcohol."



A magnificent man and his flying machine

AUGUST 1848

ABORIGINAL BOOMERANG—"The Boomering is a curious instrument used as an offensive weapon by the blacks of Australia, and in their hands, it performs most wonderful and magic actions. A late resident of that strange country published a description of some of the feats performed by the Boomering. The instrument itself is a thin curved piece of wood up to three feet in length and two inches broad—one side is slightly rounded, the other quite flat. An Australian black can throw this whimsical weapon so as to cause it to describe a complete circle in the air; the whole circumference of the circle is frequently not less than two hundred and fifty yards."

NEWS AND ANALYSIS

IN FOCUS

DEADLY SECRETS

As India and Pakistan demonstrated, military nuclear know-how is spreading with frightening ease

Although observers have presumed for years that both India and Pakistan had nuclear arsenals, the 11 nuclear detonations on the Indian subcontinent in May heralded a new level of tension between foes that have fought three conventional wars in their 51-year history as independent nations. More broadly, however, the development is illustrative of the challenges facing the international community as it seeks to halt global nuclear proliferation. India and Pakistan, which took widely divergent routes to their bombs, have shown how hard it is to deny simple but effective nuclear weaponry even to relatively poor countries, how easy it is to rend the fragile de facto moratorium on the use of these munitions and how difficult it becomes to help a developing nation build nuclear reactors while ensuring that none of the aid is put to military use.

B. K. BANGASH *AP Photo*

"WE WILL DESTROY INDIA," reads the headline on this Pakistani newspaper, after the country's nuclear tests.

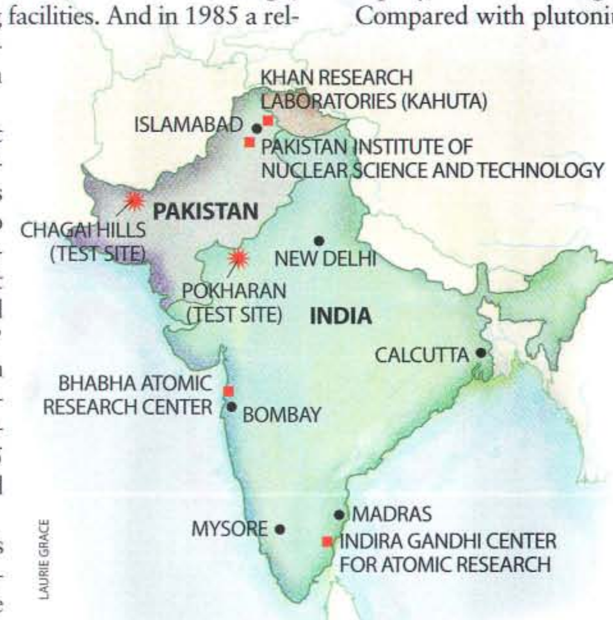
The Indian and Pakistani tests should not have been a surprise. They have roots, in fact, extending back to the early years of the nuclear era. In the 1950s, 1960s and into the 1970s, both India and Pakistan participated in nuclear cooperation efforts, most notably the Atoms for Peace program begun by the U.S. Under this program, the U.S. provided critical blueprints, know-how and components for a plant at the Bhabha Atomic Research Center in Trombay, just north of Bombay, for reprocessing the spent fuel from nuclear reactors.

Reprocessing uses a series of complex industrial-chemical processes to remove from spent fuel variant forms of elements, called isotopes, that can be reused in reactors. Essentially the same processes, though, can be used to extract the plutonium isotope 239—the key ingredient in the core of the most common types of nuclear bombs. In fact, the Trombay reprocessing plant, which was built in the late 1950s and early 1960s, was based on a process called Purex, originally developed in the U.S. to separate weapons plutonium.

Before nuclear materials can be reprocessed, they must first be irradiated in a reactor to create plutonium. And here, too, India got assistance from the West. In the late 1950s Canada provided a research reactor called Cirus. Significantly, the

reactor was a type that used so-called heavy water (enriched in the hydrogen isotope deuterium) rather than ordinary (or "light") water to lessen the energy of neutrons in the reactor's core. According to John C. Courtney, professor emeritus of nuclear engineering at Louisiana State University, this feature suited it to producing bomb-grade plutonium 239 from uranium 238, which is abundant in easily produced uranium oxide fuel. Using the Cirus reactor and the Trombay facility, Indian scientists and engineers accumulated enough plutonium to build an atomic bomb, which was detonated in 1974.

After that test, India lost all external support for its nuclear program, leaving the country on its own, for the most part, in its nuclear efforts. Indian engineers and scientists built their own heavy-water reactors, modeled on the Canadian design, as well as their own reprocessing facilities. And in 1985 a relatively large plutonium-producing reactor called Dhruva began operating at Trombay.



NUCLEAR TEST SITES AND FACILITIES enabled India and Pakistan to create the bomb.

The Indian program did not stop there. According to proliferation experts, the country has also built a plant near Mysore to produce tritium—used in thermonuclear reactions (which exploit the nuclear fusion of tritium and deuterium) and in "boosting" the yields of less powerful fission bombs. India also reportedly constructed a plant to produce uranium highly enriched in the 235 isotope, which can also be used in concocting nuclear devices.

But the focus of the country's nuclear program rests on plutonium. Peter L. Heydemann, the well-connected former science attaché at the U.S. embassy in New Delhi, estimates that by using Dhruva and other reactors India accumulated some 1.2 tons of plutonium, which could yield dozens if not hundreds of bombs, depending on the program's technical sophistication. With lavish government funding and a large pool of top scientific and engineering talent, India's nuclear establishment swelled in the 1980s to become the only one in the developing world that encompasses all aspects of what is known as the nuclear fuel cycle, from the mining of uranium ores to the construction of reactors.

Although its nuclear talent pool is by all accounts large and impressive, India's claim that one of the three nuclear devices tested on May 11 was a true thermonuclear device was met with skepticism. Such a device exploits nuclear fusion to achieve explosive yields that can be thousands of times greater than those of fission devices. Indian officials said that the purported thermonuclear device had an explosive yield of 43 kilotons. But outside experts, using seismic data, have estimated that the yield may have been as low as 25 kilotons. By thermonuclear standards, these values are rather small.

Given the history of deceit in nuclear weapons programs, many observers have suggested that the Indians tested a boosted-fission device rather than a true thermonuclear bomb or that they tested a thermonuclear weapon that turned out to be a dud. Yet Theodore B. Taylor, a former thermonuclear weapons designer at Los Alamos National Laboratory, says it is possible the Indians opted for a low yield in part because "it would be a more severe test. It's harder to be accurate in predicting the behavior of a smaller weapon." He also notes that the U.S. has produced tactical thermonuclear weapons in the past with yields in the tens of kilotons.

In contrast to India's heavy reliance on plutonium-based weapons, Pakistan's nuclear program is—for now—built entirely around the use of uranium 235. The isotope, which accounts for 0.72 percent of natural uranium, is separated from the useless 238 form in high-performance centrifuges at the Kahuta research center near Islamabad. The man now hailed as the father of Pakistan's bomb, Abdul Qadeer Khan, smuggled plans for the centrifuges in the 1970s from a Dutch company, Ultra-Centrifuge Nederland.

Compared with plutonium, uranium has a key disadvantage as a bomb material: the critical mass needed is substantially larger, making it harder to fashion into warheads that can be launched on missiles. U.S. intelligence indicates that Pakistan's bombs were relatively simple Chinese pure-fission designs. Such bombs would need more than 15 kilograms of weapons-grade uranium (material containing at least 93 percent of the 235 isotope). Plutonium-based bombs, in contrast, can make do with as little as three to six kilograms of fissile material. The Institute for Science and International Security (ISIS) in Washington, D.C., estimates that Pakistan could by now have produced enough weapons-grade uranium at Kahuta for up to 29 bombs, with the total growing by five each year. Pakistan apparently chose uranium because during the 1970s international safeguards thwarted its attempts to separate plutonium covertly from irradiated fuel. This past April, however, Pakistan said it had commissioned a new reactor at Khushab in the province of Punjab that is not subject to safeguards. According to ISIS, Pakistan should within a few years be able to produce 10 to 15 kilograms of plutonium a year. The country already has reprocessing capability, so it could in the future supplement its uranium bombs with lighter plutonium ones.

Diplomatic efforts to avert a catastrophe in the Indian subcontinent will now focus on persuading both of the newly declared nuclear powers to refrain from manufacturing warheads and especially from placing them on missiles, which could strike each other's cities within minutes. Missile technology is well advanced in the region—thanks in part to U.S. supercomputers that IBM and Digital Equipment Corporation supplied to India, according to Gary Milhollin of the Wisconsin Project on Nuclear Arms Control. The best possible outcome, Milhollin says, would be one in which India and Pakistan are persuaded to join the Nuclear Non-Proliferation Treaty and give up the bomb. But in the near future, the chances are not great. The world should hope both belligerents have established effective controls on their deadly new toys.

—Glenn Zorpette, with Tim Beardsley in Washington, D.C.

SCIENCE AND THE CITIZEN

PHYSICS

A MASSIVE DISCOVERY

The weight of neutrinos offers clues to stars, galaxies and the fate of the universe

Future historians may look back on 1998 as the year that particle physics got interesting again. For decades, the search for the fundamental nature of matter has been reduced to a jigsaw puzzle. The Standard Model of particle physics provided the frame, with outlines of each of the two dozen elementary particles sketched in their proper places. When an army of almost 1,000 physicists discovered the top quark in 1995, the puzzle seemed to be complete. Only a bit of bookkeeping remained: to confirm that the three lightest particles—the electron-, muon- and tau-neutrinos—indeed weigh exactly nothing, as the Standard Model predicts.

But in June the 120 Japanese and American physicists of the Super-Kamiokande Collaboration presented strong evidence that at least one of the neutrinos (and probably all of them) weighs something. That neutrinos have a small mass is no small matter. It could help explain how our sun shines, how other stars explode into brilliant supernovae and why galaxies cluster in the patterns that they do. Perhaps most important, explains Lincoln Wolfenstein, a physicist at Carnegie Mellon University, "once you accept that one neutrino has mass, you realize that the truth is something beyond the Standard Model."

Nearly all neutrino physicists have accepted the conclusion, because the new data are supported by several years of similar observations at other detectors. John N. Bahcall of the Institute for Advanced Study in Princeton, N.J., says the evidence "seems completely convincing to me. It is simply beautiful!"

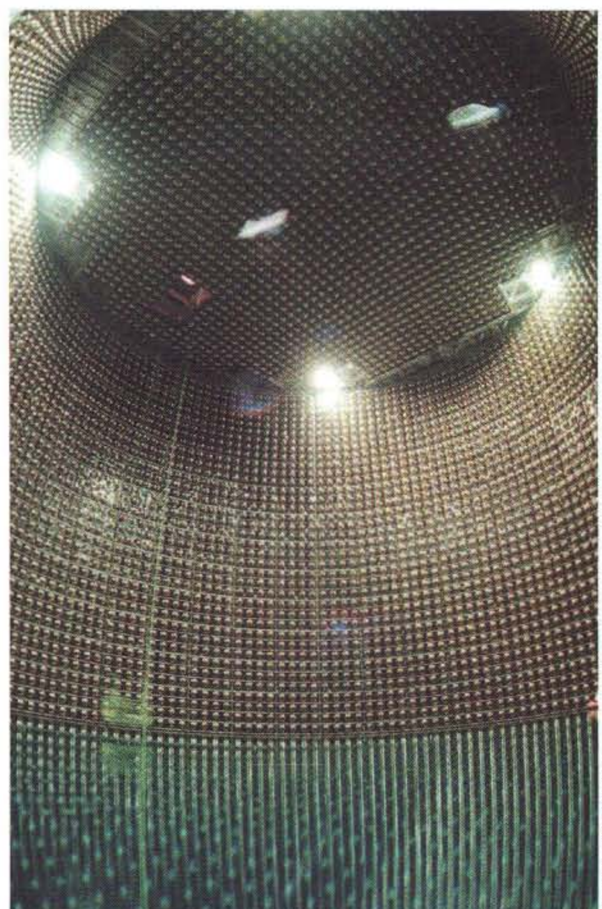
The neutrino is certainly sublime in its subtlety. Quarks and electrons are impossible to miss; we and our world are made from them. The muon and tau, cousins of the electron, are unfamiliar because they die almost at birth. But neutrinos surround us perpetually, yet invisibly. Trillions zip through your body as you read this. Created by the big bang, by stars and by the collision of cosmic rays with the earth's atmosphere, neutrinos outnumber electrons and protons by 600 million to one. "If they have a mass of just one tenth of an electron volt [an electron, in comparison, weighs about 500,000 eV], then neutrinos would account for about as much mass as the entire visible universe," says Joel R. Primack, a cosmologist at the University of California at Santa Cruz.

About 0.1 eV now seems to many physicists a likely mass for the muon-neutrino. They can't be certain yet, because the only way to weigh particles that can zoom almost unhindered through the earth at nearly the speed of light is to do so indirectly. By patiently watching a high-tech cistern buried 2,000 feet underneath a Japanese mountain, physicists working on the Super-Kamiokande project could record faint flashes emitted on the exceedingly rare occasions when a muon- or electron-neutrino collided with one out of the 50,000 tons of water molecules in the tank. Over time, traces from those neutrinos that had been created in the atmosphere started to reveal a pattern. Those arriving from above came in the expected proportion and number. "We even saw a hot spot toward the east caused by a well-known asymmetry in the earth's magnetic field" that creates more cosmic-ray collisions in that direction, says Todd J. Haines of Los Alamos National Laboratory. But too few muon-neutrinos arrived from below.

Two large groups of physicists worked independently to explain why. Both ruled out all explanations save one: the three kinds of neutrinos are not different particles in the way that electrons and muons are. Each neutrino is in fact a mixture of three mass states. The mixture can change as the neutrino travels, transforming muon-neutrinos created above South America into heavier tau-neutrinos by the time they reach the detector in Japan. That is why too few muon-neutrinos appeared in the Super-Kamiokande tank; some had metamorphosed into another, undetectable type.

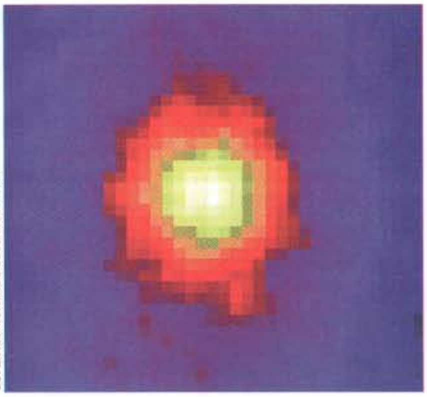
Theorists figure that there is no way to have mass states without having mass. But so far all that Henry W. Sobel, a physicist at the University of California at Irvine and spokesman for the collaboration, can say is that the difference between the mass of muon-neutrinos and whatever they are changing into is between 0.1 and 0.01 eV—definitely not zero.

Wolfenstein points out that "this does not solve the solar-neutrino problems," the most baffling of which is the fact that only half the electron-neutrinos that



KYODO NEWS INTERNATIONAL, INC.

PHOTOMULTIPLIER TUBES lining the Super-Kamiokande detector record the collision of neutrinos with water molecules.



SUPER-KAMIOKANDE COLLABORATION

SUN SEEN IN "NEUTRINO LIGHT" should theoretically shine brighter than it does.

theoretically should fall from the sun to the earth are actually detected here. But Bahcall adds that it does "strengthen the conviction of nearly everyone involved in the subject that the explanation of the solar-neutrino problems is oscillations" of neutrinos from one variety to another on their journey to the earth.

"A little hot dark matter in the form of massive neutrinos may be just what is needed" to help reconcile another astrophysical accounting discrepancy, Primack says. Many lines of evidence suggest that there is about 10 times more matter in the universe than human instruments can see. Neutrinos will now fill in some of that missing matter.

Whether neutrinos weigh enough to make a significant difference in the fate of the universe and the composition of matter remains a mystery. "We can't build theories on this without firmer data about the masses and transition amplitudes," says Steven Weinberg, one of the architects of the Standard Model and a professor at the University of Texas. "We're still far from that. But there are some very important experiments in the wings that may answer those questions."

In January physicists will create a beam of muon-neutrinos in an accelerator near Tokyo and aim it at the Super-Kamiokande detector. Within a few years, scientists at Fermi National Accelerator Laboratory in Batavia, Ill., hope to send swarms of the particles flitting toward a detector deep in a Minnesota mine shaft. These controlled experiments may finally fill the last holes in the Standard Model and, if we are lucky, reveal it to be only a part of a much larger, more beautiful picture.

—W. Wayt Gibbs in San Francisco

SCIENCE AND RELIGION

BEYOND PHYSICS

Renowned scientists contemplate the evidence for God

Modern science, like every successful philosophy, has axioms that it takes on faith to be true. Allan R. Sandage, one of the fathers of modern astronomy, has just slid one of these precepts onto an overhead projector. In letters too large to ignore, it hangs before the eyes of several hundred scientists, theologians and others gathered here at the University of California at Berkeley to discuss the points of conflict and convergence between science and religion. The axiom is called Clifford's dictum: "It is wrong always, everywhere and for everyone to believe anything on insufficient evidence."

Is there sufficient evidence to support a belief in a Judeo-Christian God? Although many scientists working in the U.S. would doubtless agree with Sandage that "you have to answer the question of what is 'sufficient' for yourself," recent polls suggest that most of them would nonetheless answer no. But the program for this conference includes some two dozen scientists, nearly all of them at the top of their fields, who have arrived at a different conclusion.

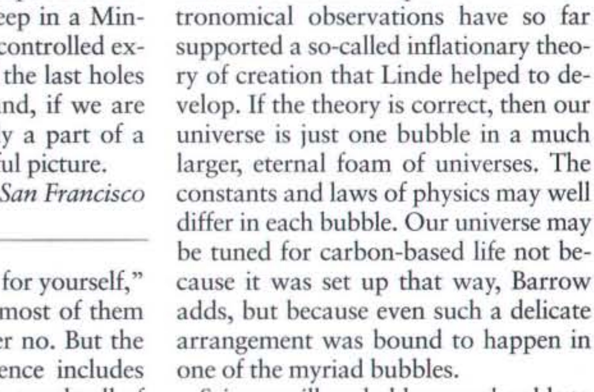
"There is a huge amount of data supporting the existence of God," asserts George Ellis, a cosmologist at the University of Cape Town and an active Quaker. "The question is how to evaluate it," he says, because the data only rarely yield to scientific analysis.

Item one on Ellis's list of evidence is the so-called Anthropic Principle. In recent decades, explains John D. Barrow, an astronomer at the University of Sussex, physicists have noticed that many of the fundamental constants of nature—from the energy levels in the carbon atom to the rate at which the universe is expanding to the four perceptible dimensions of space-time—are just right for life. Tweak any of them more than a tad, and life as we know it could not exist.

John Polkinghorne, a particle physicist turned Anglican priest, points out other curiosities. "How is it that humans' cognitive abilities greatly exceed the demands imposed by evolutionary pressures, so that we can perceive the quantum nature of the universe and map its cosmic features?" he asks. And why is mathematics so surprisingly effective at describing the physical world? One possible explanation, Polkinghorne, Ellis and other well-respected physicists argue, is that the universe was designed.

"Certainly more and more top-level scientists are considering the Anthropic Principle seriously in their work," concedes Andrei D. Linde, a Stanford University cosmologist. But he disputes that the coincidences point to God. Astronomical observations have so far supported a so-called inflationary theory of creation that Linde helped to develop. If the theory is correct, then our universe is just one bubble in a much larger, eternal foam of universes. The constants and laws of physics may well differ in each bubble. Our universe may be tuned for carbon-based life not because it was set up that way, Barrow adds, but because even such a delicate arrangement was bound to happen in one of the myriad bubbles.

Science will probably never be able to determine which of the cases is true. "We are hitting the boundaries of what is ever going to be testable," Ellis says. And not just in cosmology. "The sci-



JOHNNY JOHNSON

***OF ALL POSSIBLE UNIVERSES,** only those similar to our own could support carbon-based life. Was ours custom-made, or are we just lucky?*

ence of the 20th century is showing us, if anything, what is unknowable using the scientific method [and] what is reserved for religious beliefs," argues Mitchell P. Marcus, chairman of computer science at the University of Pennsylvania. "In mathematics and information theory, we can now *guarantee* that there are truths out there that we cannot find."

"The inability of science to provide a basis for meaning, purpose, value and ethics is evidence of the necessity of religion," Sandage says—evidence strong enough to persuade him to give up his atheism late in life. Ellis, who similarly turned to religion only after he was well established in science, raises other mysteries that cannot be solved by logic alone: "the reasons for the existence of the universe, the existence of any physical laws at all and the nature of the physical laws that do hold—science takes all of these for granted, and so it cannot investigate them."

"Religion is very important for answering these questions," Sandage concludes. But how exactly? Pressing the scientists on the details of their beliefs reveals that most have carved a few core principles out of one of the major religious traditions and discarded the rest. "When you start pushing on the dogma, most scientists tend to part company," observes Henry S. Thompson of the University of Edinburgh.

Indeed, for Ellis, "religion is an experimental endeavor just like science: all doctrine is a model to be tested, and no proof is possible." Sandage confesses that, like many other theoretical physicists, "I am a Platonist," believing the equations of fundamental physics are all that is real and that "we see only shadows on the wall." And Pauline M. Rudd, a biochemist at the University of Oxford, observes: "I have experiences that cannot be expressed in any language other than that of religion. Whether the myths are historically true or false is not so important."

There seem only two points on which all the religious scientists agree. That God exists. And that, as Albert Einstein once put it, "science without religion is lame; religion without science is blind."

—*W. Wayt Gibbs in Berkeley, Calif.*

PUBLIC HEALTH

FOOD FOR THOUGHT

Cutbacks in the food stamp program are proving hazardous to the nation's health

If diabetics don't eat, they stop taking insulin and can develop ketoacidosis, a potentially life-threatening complication of diabetes that can be severe enough to send someone into a coma. So when Nicole Lurie noticed that many of her diabetic patients were coming to Hennepin County Medical Center in Minneapolis with ketoacidosis because they could not afford to eat, she was understandably alarmed. Lurie, a professor of medicine and public health at the University of Minnesota, along with her colleagues Karin Nelson and Margaret E. Brown, decided to investigate just how prevalent hunger was among all their patients at the hospital. Now Lurie describes the diabetics as "canaries in the coal mine," providing advance warning of hunger's widespread threat to health.

The Minnesota team reported its findings in a recent issue of the *Journal of the American Medical Association:* of the 567 patients interviewed in early 1997, 13 percent reported that during the previous year they had, on several occasions, not eaten for an entire day, because they could not afford food. A separate survey of 170 diabetics revealed that almost 19 percent had suffered complications that resulted from not having enough money to eat. Hennepin County Medical Center, which is a public hospital in an urban setting, treats many low-income patients. Indeed, Lurie and her colleagues note that among patients reporting hunger or uncertainty about when they might eat next, many had annual incomes of less than $10,000. Many also shared another characteristic—recent reductions in their food stamp benefits.

The 1996 welfare reform bill trimmed $27 billion over six years from the food stamp program, which is run by the U.S. Department of Agriculture (USDA). The program provides food vouchers to people with gross incomes of less than 130 percent of federal poverty guidelines. (That is, the income of a family of three must be less than around $1,450 per month.) Nearly all recipients have felt the effects of the cutbacks, but the reduction hit two groups particularly hard. The new law denied food stamps to all legal immigrants (illegal immigrants have never been eligible). And it put new financial pressure on unemployed recipients, by limiting the maximum duration of benefits. Now unemployed able-bodied adults under 50 with no dependents may receive food stamps for no longer than three months during any three-year period.

Congress recently voted to restore food stamp benefits to certain categories of legal immigrants—such as children, the elderly and the disabled—after researchers documented that these groups were being harmed by the cuts. One disturbing study by Physicians for Human Rights, for example, found that 79 percent of legal immigrant households surveyed in March were "food insecure"—meaning they often went hungry and worried about when their next meal would be. The study coordinator, Jennifer Kasper of the Boston Universi-


B. DAEMMRICH *Image Works*

***FEEDING THE HUNGRY** often falls to food banks, such as this branch of the Salvation Army in San Antonio, Tex.*

IN BRIEF

New Planet?
A California astronomer has snapped what appears to be the first image of a planet outside our solar system. Using the Hubble Space Telescope, Susan Terebey photographed the object, believed to be two to three times the mass of Jupiter, escaping from a pair of young binary stars 450 light-years away. The supposed planet is connected to the stars by a filament of light, probably an artifact of its trajectory. Life on the new body is unlikely, as its surface temperature is several thousands of degrees. Even so, the discovery could change ideas of how planets usually form.

NASA AND SUSAN TEREBEY *Extrasolar Research Corporation*

Polarized Vision
Researchers have figured out why squid don't squint: like several other color-blind animals (but unlike people), these leggy ocean dwellers are visually sensitive to polarized light. Roger T. Hanlon, director of the Marine Resources Center at the Marine Biological Laboratory in Woods Hole, Mass., and his colleagues determined that polarization helps to enhance the contrast of the squid's black-and-white vision—enabling it to better detect prey, such as plankton, that have evolved transparent bodies for protection from predators.

Nintendo Neurology
Video games play with your brain. Paul M. Grasby and his colleagues at Hammersmith Hospital in London gave a drug called raclopride, which binds to receptors for the neurotransmitter dopamine, to eight male volunteers. The drug was tagged with low-level radiation. Next, they monitored brain activity using positron emission tomography (PET) as the subjects played video games or stared at blank screens. During the games, raclopride binding decreased in the striatum, indicating a surge of dopamine secretion there. Moreover, the increase in dopamine was as large as that seen when subjects are injected with amphetamines or the stimulant Ritalin.

More "In Brief" on page 14

ty School of Medicine, notes that because hunger has been linked to asthma, diarrhea and anemia in children, "policymakers should consider the true cost of cutting food stamps."

Lurie agrees. "Cuts in food stamps may not be benign," she says. "Squeeze food stamps, and people can end up in the hospital." She also emphasizes that physicians should talk to their patients about hunger. "Doctors need to understand when patients are making tradeoffs, buying food instead of medicine. People don't volunteer this information—doctors have to ask," she states.

When Lurie and her colleagues asked, they found that 8 percent of their diabetic patients had decreased or stopped taking their insulin because they had not eaten enough to need the normal dose. Getting enough food and enough medicine can be a challenge not just for diabetics: a survey by Second Harvest, the nation's largest charitable hunger-relief organization, found that 28 percent of people visiting food banks have at some point been forced to choose between getting medical care and buying food.

Researchers say it is notable that these alarming new findings were obtained within two years after the 1996 welfare bill was passed. As the law stands now, many of the cuts in food stamps will continue for another four years. J. Larry Brown, director of the Tufts University Center on Hunger, Poverty and Nutrition Policy, says the USDA has only recently started monitoring hunger and food insecurity more systematically.

Because it takes a couple of years to obtain a clear picture of hunger across the nation, Brown explains that he and others will be conducting smaller-scale surveys. Results should come in within several months, "enabling us to show trends faster," he notes. Brown points out the importance of identifying hunger and food insecurity before people wind up in the hospital, as they did in the Minnesota study. "By the time [hunger] shows up clinically, these people have been without food for a long time," he says. "I can separate my human side from my scientific side, and on the human side the situation is very frightening." —*Sasha Nemecek*

ANTI GRAVITY

Hoop Genes

You know how close I came to playing professional basketball? About 17 inches. Seventeen inches taller, and I would have been an even seven feet, which at least would have given me a shot at seeing just how close a shave Michael gets on the top of his head. Having immense strength, acrobatic agility, catlike coordination, unbridled desire and a soft touch on my fall-away jumper would also have come in handy, but the height thing couldn't have hurt. All of which brings us to genes.

Although the Human Genome Project will probably fail to uncover a DNA sequence governing three-point shooting, British researchers have indeed found a jock gene. The gene in question, which comes in two forms called *I* and *D* (for "insertion" and "deletion"), is for angiotensin-converting enzyme (ACE), a key player in modulating salt and water balance, blood vessel dilation and maybe more. People carrying the *D* form are perfectly normal but will probably be at home watching television reports of people who have the *I* form planting flags on the top of Mount Everest.

"We got into this system because it was thought to control heart growth," says Hugh E. Montgomery of the University College London Center for Cardiovascular Research, who is the lead author of the ACE gene study, which appeared in the May 21 issue of *Nature*. "We started looking at exercising athletes purely because their hearts grow when they exercise." Assuming that a crucial factor in endurance would be efficient use of oxygen, Montgomery and his colleagues tested "extreme endurance athletes exposed to low oxygen concentrations," namely mountaineers, all of whom were also British. (Why did the British researchers test British mountain climbers? Because they were there.)

Montgomery's group found a disproportionately high representation of the *I* form among elite mountaineers, some of whom had repeatedly climbed to 8,000 meters (over 26,000 feet) without supplemental oxygen. In another study, they followed 78 army recruits through a 10-week training program. Those with *I* and *D* forms tested equally before the workouts began, but soon the *I*s had it. All the exercisers increased their endurance, as measured by curls of a 15-kilogram (33-pound) barbell, but *I* folks improved 11 times more than the *D*s.

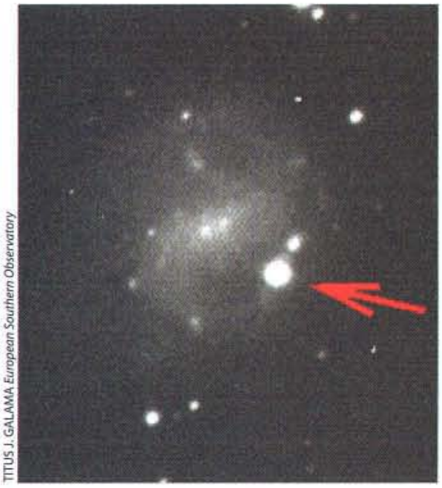
The studies bring to mind the eerie case of Eero Maentyranta, winner of three cross-country skiing gold medals at the 1964 Winter Olympics in Innsbruck. In 1993 re-

ASTRONOMY

BRIGHT LIGHTS, BIG MYSTERY

The brightest flash ever seen has revealed exotic celestial objects

Hardly an astronomical announcement makes the front pages without being said to overturn all existing theories. Gamma-ray bursts are one of the few things for which this has actually been true. These seemingly random flashes—which, if you had gamma-ray vision and didn't blink at the wrong time, would outshine the rest of the sky—were long thought to originate in our galaxy. But that theory foundered in 1991, when the Compton Gamma Ray Observatory satellite detected bursts all over the sky, not only in the Milky Way. Another hypothesis, in which bursts occur just outside our galaxy, crumbled last year when the first distance measurement of a burst put it too far away. Although astronomers now agree that the bursts are some kind of megaexplosion in distant galaxies, they still appear on most top-10 lists of cosmic mysteries.



TITUS J. GALAMA *European Southern Observatory*

UNUSUAL SUPERNOVA became visible shortly after a gamma-ray burst this spring, linking the two phenomena.

This past spring astronomers reported the distances to two new bursts—which, at first glance, seemed to scuttle some of the few remaining plausible explanations. The first of these bursts, spotted last December 14 by the BeppoSAX satellite, occurred in a galaxy 12 billion light-years away, according to observations by Shrinivas R. Kulkarni and S. George Djorgovski of the California Institute of Technology. To be so bright at such a distance, the burst must have shone more brilliantly than any object previously recorded.

Just as astronomers were reconciling themselves to the unexpected distance and brilliance of this burst, along came another on April 25 that was unexpectedly near and dim: nearly 100 times closer and 100,000 times dimmer than the December event. Even stranger, this burst was followed not by the usual afterglow of less energetic radiation but by a supernova—the first time an exploding star and gamma burst have been seen together. The supernova appeared in visible-light images made by Titus J. Galama of the University of Amsterdam and in radio observations by Mark H. Wieringa of the Australia Telescope National Facility. This was no normal supernova, either. It dimmed more slowly than others with its spectral characteristics, and it spewed out more radio power than any supernova seen before.

Amid these events, a third burst went off—the burst of hype. A widely quoted National Aeronautics and Space Administration press release called the December event "the most powerful explosion since the creation of the universe," which slighted any upheavals that release energy in forms we cannot yet detect. The hyperbole also made gamma bursts out to be something radically different from supernovae, when in fact both probably entail the death of a star and the conversion of much of its mass into pure energy.

In a supernova, the energy conversion involves the thermonuclear detonation of a white dwarf star or the implosion of a massive star. But such cataclysms could not produce gamma bursts: the stellar debris would choke off the gamma rays. Bursts must come from low-density sources. And yet their energy must ultimately originate in high-density volumes 100 kilometers across at most, or else the bursts would not flash and flicker as they do. These two requirements imply that the production

searchers found that Maentyranta and a lot of his family have a genetic mutation affecting red blood cell production. As a result, Maentyranta's blood carries up to 50 percent more hemoglobin than an average male's. And whereas having, say, a third lung might get you disqualified from competition, a set of genes that grants you access to half again as much oxygen as the competition is still legal.

Biology's newfound ability to spot individual genes associated with athleticism is intriguing but, on contemplation, not surprising. "Athletics, by already selecting for extreme phenotypes, must be selecting for a significant genetic influence," says Philip Reilly, executive director of the Shriver Center for Mental Retardation in Waltham, Mass., a clinical geneticist and an attorney who has spent the past 25 years pondering genetics and its implications for society. "In one sense, this is old news," he notes. "There is no physical trait for which we have better evidence of substantial genetic contribution to ultimate expression of phenotype than height." And for most athletes, jockeys being the ridiculously obvious exception, being big is advantageous. (The late, great sportswriter Shirley Povich actually bemoaned all the tallness. "[Basketball] lost this particular patron back when it went vertical and put the accent on carnival freaks," he wrote. "They don't shoot baskets anymore, they stuff them, like taxidermists.")

So genes for robust physical traits help make for a good athlete. But the basis for behavior, which also contributes to athleticism, remains more mysterious. Another 45 army recruits, with the same proportion of *I* and *D* forms as the 78 who trained for 10 weeks, dropped out entirely to go home. "Sometimes," Montgomery explains, "they just miss their mums." *—Steve Mirsky*

MICHAEL CRAWFORD

In Brief, continued from page 12

Spinach Power

Popeye had better join the bomb squad. Scientists at the Department of Energy's Pacific Northwest National Laboratory have found that spinach enzymes can neutralize dangerous explosives, such as TNT. Nitroreductase, found in sundry leafy greens, can eat, digest and transform these compounds into less toxic by-products, which can then be used by industry or be further reduced to carbon dioxide and water.

DAVIES & STARR, INC. *Liaison International*

Hot Finding on Mars

Philip R. Christensen of Arizona State University made a startling announcement at the American Geophysical Union's spring meeting in Boston: data from the Thermal Emission Spectrometer (TES) on board the Mars Global Surveyor—a spacecraft now in orbit around the Red Planet—indicated the presence of a large deposit of coarse-grained hematite. This rock is formed by thermal activity or bodies of water, and so geologists take the TES finding as further evidence that Mars was once warmer and may have supported life. They hope to study the deposit—which measures some 300 miles in diameter—more closely during the 2001 Lander mission.

Fast Extinction

At the end of the Permian period, more than 85 percent of all ocean species and 70 percent of terrestrial vertebrate genera died out. A study in *Science* on May 15 shows that this vanishing act happened fast—in less than a million years. Douglas H. Erwin of the National Museum of Natural History and his co-workers examined 172 samples of volcanic ash from various levels at three locations in southern China and one in Texas. The samples were readily dated because they contained zircon. This mineral incorporates uranium into its crystal lattice when it forms; over time, the element decays into lead and tells the crystal's age. Erwin found that most species had disappeared between 252.3 million and 251.4 million years ago. Although the extinction's cause is still unclear, its speed does rule out an earlier theory implicating plate tectonics. (See July 1996, page 56.)

More "In Brief" on page 16

of energy and the generation of gammas occur in separate locations.

The second of these steps has been explained by Martin J. Rees of the University of Cambridge and Peter I. Mészáros of Pennsylvania State University. A jet of radiation and electrons squirting out at near light speed would outrun any interfering debris. After this fireball had ballooned to several hundred million kilometers in size, shock waves within it would give off the gamma flash; later, collisions with surrounding matter could emit an afterglow. Radio observations by Dale A. Frail of the National Radio Astronomy Observatory have supported this explanation.

But what would produce enough energy to create such a jet? In one scenario, proposed in 1986 by Bohdan Paczyński of Princeton University, two neutron stars or a neutron star and a black hole orbit ever more tightly and eventually unite in a passionate but ruinous embrace. Stan E. Woosley of the University of California at Santa Cruz suggested another possibility in 1993. Perhaps "hypernovae," also called "collapsars"—souped-up supernovae that occur when stars are too massive to undergo normal supernovae—power the cosmic flashes.

Despite some observers' initial claims, the brightness of the bursts may not be enough to decide between these models. Both lead to the same kind of object: a black hole surrounded by a ring of debris. And both leave the same questions unanswered: What triggers the fireball? Is the radiation focused into beams or emitted uniformly? Pinpointing the bursts' locations within their galaxies may supply clues. Neutron stars wander far from their places of birth before merging, whereas the stars that undergo hypernovae stay put. So if bursts tend to occur in star-forming regions—as a few shreds of evidence now suggest—hypernovae seem the more likely source.

But there is a third possibility. The diversity of these latest bursts, Woosley says, suggests that each model accounts for some bursts. This time, rather than overturning all the theories, observers may have done the opposite: confirmed them, all of them. *—George Musser*

PROFILE

An Express Route to the Genome?

In his race to beat the Human Genome Project, J. Craig Venter has riled geneticists everywhere

J. Craig Venter, the voluble director of the Institute for Genomic Research (TIGR) in Rockville, Md., is much in demand these days. A tireless self-promoter, Venter set off shock waves in the world of human genetics in May by announcing, via the front page of the *New York Times*, a privately funded \$300-million, three-year initiative to determine the sequence of almost all the three billion chemical units that make up human DNA, otherwise known as the genome. The audacious claim prompted incredulous responses from mainstream scientists engaged in the international Human Genome Project, which was started in 1990 and aims to learn the complete sequence by the year 2005. This publicly funded effort would cost about 10 times as much as Venter's scheme. But Venter's credentials mean that genome scientists have to take his plan seriously.

In 1995 Venter surprised geneticists by publishing the first complete DNA sequence of a free-living organism, the bacterium *Haemophilus influenzae*, which can cause meningitis and deafness. This achievement made use of a then novel technique known as whole-genome shotgun cloning and "changed all the concepts" in the field, Venter declares: "You could see the power of having 100 percent of every gene. It's going to be the future of biology and medicine and our species." He followed up over the next two and a half years with complete or partial DNA sequences of several more microbes, including agents that cause Lyme disease, stomach ulcers and malaria.

The new private human genome initiative will be conducted by a company (yet to be named) that will be owned by

TIGR, Perkin-Elmer Corporation (the leading manufacturer of DNA sequencers) and Venter himself, who will be its president. He plans to warm up for the human sequence next year by knocking off the genome of the fruit fly *Drosophila melanogaster,* an organism used widely for research in genetics.

JEAN-CHRISTIAN BOURCART *Gamma Liaison*

AHEAD OF THE PACK:
J. Craig Venter admires a model of his racing yacht.

Venter, 51, has a history of lurching into controversy. As an employee of the National Institutes of Health in the early 1990s, he became embroiled in a dispute over an ultimately unsuccessful attempt by the agency to patent hundreds of partial human gene sequences. Venter had uncovered the partial sequences, which he called expressed sequence tags (ESTs), with a technique he developed in his NIH laboratory for identifying active genes in hard-to-interpret DNA. "The realization I had was that each of our cells can do that better than the best supercomputers can," Venter states.

Many prominent scientists, including the head of the NIH's human genome program at the time, James D. Watson, opposed the attempt to patent ESTs, saying it could imperil cooperation among researchers. (Venter says the NIH talked him into seeking the patents only with difficulty.) And at a congressional hearing, Watson memorably described Venter's automated gene-hunting technique as something that could be "run by monkeys." An NIH colleague of Venter's responded later by publicly donning a monkey suit.

Venter left the NIH in 1992 feeling that he was being treated "like a pariah." And he does not conceal his irritation that his peers were slow to recognize the merits of his proposal to sequence *H. influenzae.* After failing to secure NIH funding for the project, Venter says he turned down several tempting invitations to head biotechnology companies before finally accepting a $70-million grant from HealthCare Investment Corporation to establish TIGR, where he continued his sequencing work. Today, when not dreaming up audacious research projects, Venter is able to relax by sailing his oceangoing yacht, the *Sorcerer.*

His assault on the human genome will employ the whole-genome shotgun cloning technique he used on *H. influenzae* and other microbes. The scheme almost seems designed to make the Human Genome Project look slow by comparison. To date, that effort has devoted most of its resources to "mapping" the genome—defining molecular landmarks that will allow sequence data to be assembled correctly. But whole-genome shotgun cloning ignores mapping. Instead it breaks up the genome into millions of overlapping random fragments, then determines part of the sequence of chemical units within each fragment. Finally, the technique employs powerful computers to piece together the resulting morass of data to re-create the sequence of the genome.

Predictably, Venter's move prompted some members of Congress to question why government funding of a genome program was needed if the job could be done with private money. Yet if the goal of the Human Genome Project is to produce a complete and reliable sequence of all human DNA, says Francis S. Collins, director of the U.S. part of the project, Venter's techniques alone cannot meet it. Researchers insist that his "cream-skimming" approach will furnish information containing thousands of gaps and errors, even though it will have short-term value. Venter accepts that there will be some gaps but expects accuracy to meet the 99.99 percent target of the existing genome program.

Shortly after Venter's proposed scheme hit the headlines, publicly funded researchers started discussing a plan to speed up their own sequencing timetable in order to provide a "rough draft" of the human genome sooner than originally planned. Collins says this proposal, which would require additional funding, would have surfaced even without the new competition. Other scientists think Venter's plan will spur the public researchers forward.

Venter has always had an iconoclastic bent. He barely graduated from high school and in the 1960s was happily surfing in southern California until he was drafted. Early hopes of training for the Olympics on the navy swim team were dashed when President Lyndon B. Johnson escalated the war in Vietnam. But Venter says he scored top marks out of 35,000 of his navy peers in an intelligence test, which enabled him to pursue an interest in medicine. He patched up casualties for five days and nights without a break in the main receiving hospital at Da Nang during the Tet offensive, and he also worked with children in a Vietnamese orphanage.

Working near so much needless death, Venter says, prompted him to pursue Third World medicine. Then, while taking premed classes at the University of California at San Diego, he was bitten by the research bug and took up biochemistry. He met his wife, Claire M. Fraser, now a TIGR board member, during a stint at the Roswell Park Cancer Institute in Buffalo, N.Y., and took

In Brief, continued from page 14

Practical Chaos

Researchers have at last drawn a credible link between chaos theory and epileptic seizures. Klaus Lehnertz and Christian E. Elger of the University of Bonn collected 68 electroencephalograms from 16 patients. They then made use of a mathematical property from chaos theory called dimension, which is essentially a measure of complexity. The pair found that 11 minutes before the onset of seizure, there is a characteristic loss of complexity in the firing of neurons near the epileptic focus in the brain. The finding may help scientists develop more sensitive instruments to detect oncoming seizures—and better ways to prevent them.

Stained-Glass Myth

Investigators have smoothed the lumps out of an old theory, finding that medieval stained glass is not thicker at the bottom because it has slowly flowed

RIEGER BERTRAND Gamma Liaison

downward. Edgar Dutra Zanotto of the Federal University of São Carlos in Brazil calculated how long it would take for a viscous liquid to flow enough to change the visible thickness of glass. His conclusion, published in May in the *American Journal of Physics,* was that cathedral glass would need a period "well beyond the age of the universe." In fact, medieval stained glass is probably thicker at the bottom because of 12th-century manufacturing methods.

Magnetars

Six years ago Robert C. Duncan of the University of Texas at Austin and Christopher A. Thompson of the University of North Carolina at Chapel Hill predicted the existence of magnetars, ultradense stars with tremendous magnetic fields. Now they have been found. In May in *Nature,* Chryssa Kouveliotou of NASA's Marshall Space Flight Center and an international team described the x-ray measurements of SGR 1806-20. The measurements showed that the object is most likely spinning at a rate characteristic of a heavy, supermagnetic star. The team estimated SGR 1806-20's magnetic field to be a whopping 800 trillion gauss—or 100 times stronger than the magnetic fields around any other known stars. *—Kristin Leutwyler*

his research group to the NIH in 1984.

His painstaking attempts to isolate and sequence genes for proteins in the brain known as receptors started to move more quickly after he volunteered his cramped laboratory as the first test site for an automated DNA sequencer made by Applied Biosystems International, now a division within Perkin-Elmer. Until then, he had sequenced just one receptor gene in more than a decade of work, so he felt he had to be "far more clever" than scientists with bigger laboratories. Venter was soon employing automated sequencers to find more genes; he then turned to testing protocols for the Human Genome Project, which was in the discussion phase.

After leaving the government and moving to TIGR, Venter entered a controversial partnership with Human Genome Sciences, a biotechnology company in Rockville established to exploit ESTs for finding genes. The relationship, never easy, foundered last year. According to Venter, William A. Haseltine, the company's chief executive, became increasingly reluctant to let him publish data promptly. Haseltine replies that he often waived delays he could have required.

The divorce from Human Genome Sciences cost TIGR $38 million in guaranteed funding. The day after the split was announced, however, TIGR started to rehabilitate itself with a suspicious scientific community by posting on its World Wide Web site data on thousands of bacterial gene sequences.

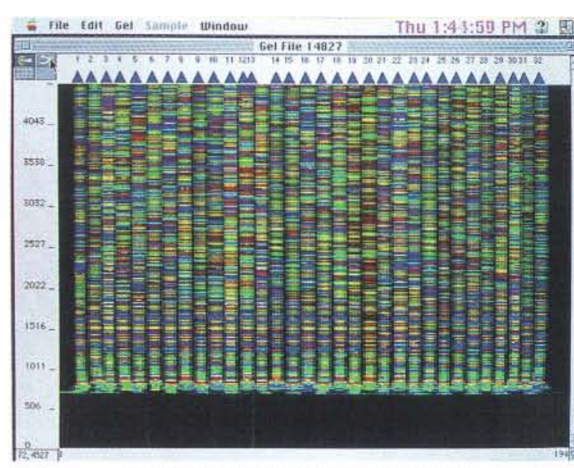
The sequencing building for Venter's new human genome company, now under construction adjacent to TIGR, should be a technological mecca. The firm will produce more DNA data than the rest of the world's output combined, employing 230 Perkin-Elmer Applied Biosystems 3700 machines, which are now in final development. These sophisticated robots, which will sell for $300,000 apiece, should require much less human intervention than state-of-the-art devices. Venter says the new venture will release all the human genome sequence data it obtains, at three-month intervals. It plans to make a profit by selling access to a database that will interpret the raw sequence data as well as crucial information on variations between individuals that in principle should allow physicians to tailor treatments to patients' individual genetic makeups. Most of the federal sequencing centers do not look at the data they produce, Venter thinks, but just put it out as if they were "making candy bars or auto parts."

The unnamed new TIGR–Perkin-Elmer operation will also patent several hundred of the interesting genes it expects to find embedded within the human genome sequence. Venter defends patents on genes, saying they pose no threat to scientific progress. Rather, he notes, they guarantee that data are available to other researchers, because patents are public documents. His new venture will not patent the human genome sequence itself, Venter states.

The plan, if it comes to pass, should be a boon for biomedical research. Yet more than 100 scientists who met recently to decide on further plans for the existing, more complete and thorough international Human Genome Project were unanimous that it, too, should go forward. Collins wonders whether the entrepreneur at TIGR will in fact be able to stick to his stated policy of releasing all data, given financial exigencies. But he and Venter have both pledged to cooperate.

Venter is unfazed. "We will make the human genome unpatentable" by placing it in the public domain, he proclaims. For the next three years, all eyes will be on Venter to see how closely he can approach that goal.

—Tim Beardsley in Washington, D.C.



PERKIN-ELMER CORPORATION

READOUT OF GENETIC DATA ***comes from a DNA sequencing machine. More advanced units will tackle the entire human genome.***

Amphibians at Risk

Some 5,000 species of amphibians inhabit the world, mostly frogs, toads and salamanders, and they seem to be dying at unprecedented rates. This situation has raised alarm because amphibians are widely regarded as uniquely sensitive indicators of the planet's health. Much of the damage to amphibians comes from habitat destruction, particularly the draining of wetlands, but what has scientists particularly worried are the declines and apparent extinctions in areas far removed from obvious human intrusion, such as the cloud forest at Monteverde, Costa Rica, where the golden toad, once abundant, has not been seen since 1989.

Do reports such as these indicate a worldwide amphibian crisis? Not necessarily, according to Joseph Pechmann of Florida International University, whose work suggests that reported declines and extinctions in near-pristine environments could simply be natural year-to-year variations: a drought, for instance, that affects egg laying and larvae survival. Because of such fluctuations, it is often impossible, in the absence of more complete historical information, to judge whether a reported decline is natural or a reaction to human activity. On the other hand, many researchers, including Andrew R. Blaustein of Oregon State University, believe the reported rates of decline and extinction are so extraordinary that they cannot be a part of the natural cycle.

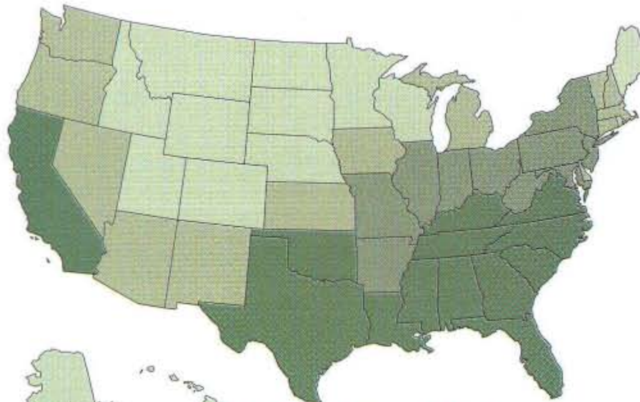
If the reported declines and extinctions are indeed highly abnormal—and this seems to be the majority point of view—what might have led to them? A variety of causes is probably at work here. Among the leading culprits are acid rain, synthetic chemicals, metallic contaminants and infectious diseases. Excessive ultraviolet radiation (presumably caused by the thinning of the ozone layer) working synergistically with fungal disease may explain some of the declines. Another possible cause is global warming–related droughts, a particular threat to amphibians, which generally require high humidity or an aquatic environment.

There has also been a recent surge of reported amphibian abnormalities, such as missing limbs and eyes. One of the factors could be a substance like retinoic acid, which may be present in water naturally or as a residue from pesticides; it produces birth defects in vertebrates. It is not clear, however, whether the increase in reports stems from greater human activity or is simply the result of more surveys: reports on amphibian abnormalities are not new and have fluctuated in number since 1700.

One of the few areas with reasonably complete information is the U.S. The Nature Conservancy and the Natural Heritage Network have identified 242 native amphibian species that have inhabited the 50 states since the beginning of European settlement. Three of these species are presumed to be extinct; an additional three are classified as possibly extinct. Moreover, 38 percent are classified as imperiled or vulnerable, a level that suggests that amphibians are more affected by human activity than birds, mammals and reptiles, which are at far lower levels of risk in the U.S. Freshwater fish are at about the same level of risk as amphibians; crayfish and freshwater mussels are at substantially higher levels of risk.

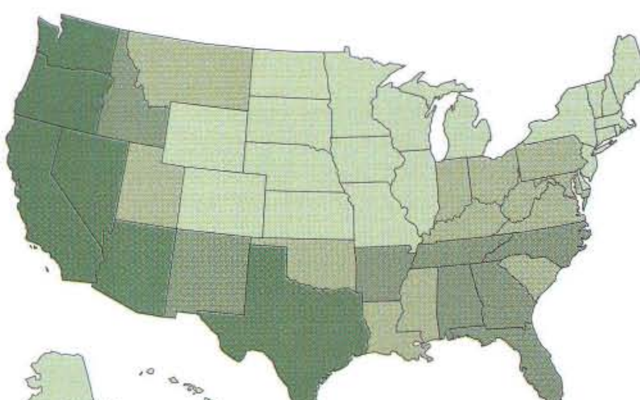
As the top map shows, the number of amphibian species in the U.S. varies considerably, with California and the South having the largest number. The bottom map shows that amphibians tend to be at greater proportional risk in the Far West. In much of the West, habitats are discontinuous, making it more difficult for locally threatened groups to be recolonized from other areas. In California's Central Valley, pesticides and herbicides are suspected of contributing to amphibian declines. Amphibians in certain parts of the West may be adversely affected by the introduction of nonnative species such as the American bullfrog, which compete with native amphibians for food, or by the introduction of salmon and trout, which eat amphibian eggs, tadpoles and even adults. The Southeast—which has more areas of continuous suitable habitat (such as the southern Appalachians) and a warm, moist climate—is better suited for recolonization.

—*Rodger Doyle (rdoyle2@aol.com)*



TOTAL NUMBER OF AMPHIBIAN SPECIES

LESS THAN 20 | 20 TO 29 | 30 TO 49 | 50 OR MORE



SPECIES AT RISK (PERCENT)

0 | 1 TO 9 | 10 TO 19 | 20 OR MORE

SOURCE: The Nature Conservancy and Natural Heritage Network in cooperation with the Association for Biodiversity Information. All data are for 1997. Excluded are amphibian species not native to their areas.

RODGER DOYLE

ENVIRONMENTAL POLICY

TRADE RULES

A World Trade Organization decision about sea turtles raises doubts about reconciling economics and the environment

The 1990s have seen the emergence of a conciliatory credo: business and environmental interests are not just compatible, they are inextricable. A healthy environment and natural-resource base are prerequisites for a healthy economy; a smoothly functioning world market, in turn, produces the resources and mind-set needed to protect the environment. Concepts such as "sustainable development" and "eco-labeling" have accordingly promised consumers a hand in making the market environmentally accountable.

Although the rhetoric seems reasonable to many on both sides of the divide, it is proving difficult to implement. As an April decision by the World Trade Organization (WTO) indicates, the interests of one nation's consumers may be consistently forced to yield to the interests of free trade. The ruling stated that a U.S. law prohibiting the import of shrimp caught in nets that can entrap sea turtles was a barrier to trade. According to the WTO, the U.S. must import shrimp from Thailand, Malaysia, India and Pakistan—regardless of whether the harvests endanger sea turtles—or face large fines.

The WTO ruling—which the U.S. says it will appeal—is not the first to subsume a nation's environmental laws or its consumers' desires to the demands of global markets. Europeans are legally required to import U.S. hormone-treated beef, Americans must stomach tuna from Mexico that endangers dolphins, and the U.S. Environmental Protection Agency lowered air-quality standards to allow imports of reformulated gasoline. The shrimp ruling, however, has mobilized the environmental community in an unprecedented fashion.

Part of the reason is that all seven species of sea turtle are endangered. Populations of these mysterious, long-lived creatures have plummeted in the past 50 years. But the other stimulus for the outcry is a growing concern over the WTO's environmental mores and the role of so-called processing and production methods. Greatly simplified, trade law states that one country cannot exclude another's product if it is "like" most such products. Therefore, the U.S. cannot ban Asian shrimp if it looks and tastes like other shrimp. The production method is not a consideration. "That is a distinction that the trade system does not want to recognize," explains Daniel A. Seligman of the Sierra Club. And it is on precisely that distinction that eco-labeling and consumer choice hinge: Was wood harvested sustainably? Did shrimpers harm an endangered species?

"Industry has increasingly been arguing that labeling is a trade barrier because it establishes different standards," Seligman continues. In other words, if rulings such as this one continue to hold, trade could lose some of its power to enact environmental change.


MIKE WEBER Center for Marine Conservation

LOGGERHEAD TURTLE WAS CAPTURED LIVE on a shrimp trawler with other bycatch. Many other sea turtles aren't so lucky.

At the same time, the WTO ruling hardly represents a black-and-white case of trade trumping environment. Just as Americans resist being force-fed a product, the Asian countries are resisting U.S. efforts to force its domestic law down their throats. Further complicating the issue is the murky history of the U.S. position: the Clinton administration finds itself defending a legal interpretation it did not countenance in the first place.

The shrimp conflict has its roots in Public Law 101-162, Section 609, which the federal government enacted in 1989. The statute stated that the U.S. would not buy shrimp caught in nets without turtle-excluder devices, or TEDs. These grills sit in the necks of shrimpers' nets, allowing small crustaceans through but stopping larger creatures and diverting them out of the trap. By 1993 the government had effectively applied the law to fishermen plying the Caribbean and western Atlantic. (U.S. shrimpers had been required to use TEDs since 1987.)

As applied, the law did little to protect sea turtles worldwide. Shrimp caught in Asia, for example, was not prohibited from U.S. markets. So several organizations, including the Earth Island Institute, challenged the government, claiming that catches of all wild shrimp come under the law's jurisdiction. In December 1995 the U.S. Court of International Trade agreed: the U.S. must embargo all shrimp caught in TED-less nets.

The government then interpreted the law to mean that foreign imports should be checked on a shipment-by-shipment basis. But the plaintiffs argued that such checking does not protect sea turtles, because TED-less boats could hand off their catch to TED-certified trawlers.

So, in October 1996, the trade court ruled against the shipment-by-shipment interpretation. (The future of the ruling is unclear, as it was recently vacated, or annulled, during appeal.) Thus, American law mandated that other countries set up and enforce regulations requiring TEDs if they wanted access to the U.S. market. Countries were given a few months to do so.

Things quickly became dicey. Thailand—which already used TEDs—joined India, Malaysia and Pakistan in opposing the ban. These nations argued that

the U.S. had no right to mandate their domestic policy and that the U.S. had not approached them to negotiate an agreement. "To me, this is perfectly reasonable for these four countries to say this is something that we do not buy and we will not do it," says Jagdish N. Bhagwati, a professor of economics at Columbia University. "The great advantage of rulings like this is that it forces the U.S. to go talk to countries and get into some discussion with the countries."

Such discussion did take place with Caribbean and Latin American nations. An Inter-American Convention on the Protection and Conservation of Sea Turtles was drafted in 1996; as of this spring, only six of the original 25 had signed it, however. "Multilateral agreement is the way to go," agrees Deborah T. Crouse, formerly of the Center for Marine Conservation. "But it is also much slower and more labor intensive. It took two years to get the Inter-American Convention, and the U.S. has not ratified or implemented it."

None of the four Asian countries contest that sea turtles are endangered. Several have conservation efforts in place, such as closing beaches to allow nesting, and have sought out TED training courses. Each has also signed the Convention on International Trade in Endangered Species (CITES), which lists all sea turtles as threatened.

But it is the muddy legal area where the WTO and CITES—as well as other environmental treaties—intersect that concerns environmentalists. CITES prohibits the trade of endangered species; the WTO prohibits barriers to trade such as objections to processing and production methods. What should happen when a production method threatens an endangered species? The WTO offers nothing but ambiguity on this front.

Formed in 1995 to replace the General Agreement on Tariffs and Trade (GATT), the WTO does make environmental provisions. The preamble to its charter mentions sustainable development; Article XX of GATT, which still holds for the WTO, offers exceptions to the rule of free trade; and the organization set up the Committee on Trade and the Environment. But that committee has not clarified the relation between the WTO and international environmental treaties.

Some lawyers point out that CITES, the Basel Convention (which prohibits the export of hazardous waste) and the Montreal Protocol (which limits chlorofluorocarbon sales) could be interpreted as illegal under the WTO: each presents barriers to trade. And the legality of consumer choice—particularly as expressed through eco-labeling—remains just as unresolved. Even given its idiosyncratic intricacies, the WTO shrimp ruling suggests that trade and the environment are still far from being happily united. *—Marguerite Holloway*

COMPUTER SCIENCE

THE *OTHER* COMPUTER PROBLEM

Will computers be ready for the euro and *the year 2000?*

In May the European Union moved one step closer to economic superpower when it officially adopted the euro as the single currency for 11 of its member nations. But as politicians celebrated, computer managers around the world issued a collective moan. They must now refit every system that deals in marks, francs, guilders and other coins of the continent to handle the new currency, which will be phased in starting in January 1999. Faced with both the euro and widespread year 2000 bugs, "a lot of companies in Europe will find themselves in dire straits," predicts Achi Racov, the chief information officer of the London-based NatWest Group. He expects that his firm will spend in excess of £200 million (about $320 million) to enable its systems to handle both the euro and the turn of the millennium.

Adding a new currency to financial systems sounds straightforward, but it is not. The European Commission in Brussels has published strict rules that computers must follow. Converting from francs to marks, for example, could soon require "triangulation," in which francs are changed first into euros, then into marks. Some software will need to display both the local currency and the euro during the transition period, from 1999 to 2002. And terabytes of historical financial data must eventually be converted into euros.

The euro will also have a widespread secondary effect on computers. No longer vulnerable to fluctuations in monetary exchange rates, companies will increase their use of cheaper supplies and workers in other countries. Such operational changes will require revised or new software. "The euro touches just about every part of British Airways from a business point of view," asserts Nigel T. Barrow, euro computer manager for the airline. All told, the cost of modifying systems in Europe (including those owned by U.S. multinational corporations) will run between $150 billion and $400 billion, states the Gartner Group, a consultancy based in Stamford, Conn.

Euro conversion projects at this time will cost more, and will progress more slowly, than they otherwise would because they will be competing for qualified programmers against the year 2000 effort. The U.S. is already short more than 100,000 information technology professionals, according to industry estimates. That number is expected to balloon in the near future.

One obvious way to solve the problem of the two largest software projects in history occurring almost simultaneously would be to push back the deadline for euro compliance from 2002 to 2005, advises Capers Jones, chairman of Software Productivity Research in Burlington, Mass. Otherwise, he claims,

YVES LOGGHE *AP Photo*

EURO—THE NEW, SINGLE CURRENCY for 11 European nations—will require extensive software modifications to financial computer systems.

more than 35 percent of the software in western Europe that needs millennium repairs will not be ready in time, leading to computer failures.

Political concerns, however, have so far outweighed the technical ones. "It's not politically feasible to push back the deadline," states Pieter Dekker of the European Commission. "It would be like trying to postpone the presidential elections in the U.S.," he concludes.

The danger, of course, is that untreated computers will crash, forcing people to enter orders and to write invoices by hand. "Unfortunately," says the Gartner Group's Nick Jones, "the politicians neither understand nor care about this huge and horrible burden they are placing on Europe's information technology industry." *—Alden M. Hayashi*

ECONOMICS

LOOK FOR THE UNION LABEL

New analysis of economic data shows that unionization could maximize productivity

After nearly a century of union-management warfare in the U.S., a series of nationwide surveys showing that union shops dominate the ranks of the country's most productive workplaces may come as a surprise. In fact, according to Lisa M. Lynch of Tufts University and Sandra E. Black of the Federal Reserve Bank of New York, economic Darwinism—the survival of the fittest championed by generations of hard-nosed tycoons—may be doing what legions of organizers could not: putting an end to autocratic bosses and regimented workplaces.

American industry has been trying to reinvent itself for more than 20 years. Management gurus have proclaimed Theories X, Y and Z, not to mention Quality Circles, Total Quality Management (TQM) and High-Output Management. Only in the past few years, however, have any solid data become available on which techniques work and which don't. Businesses do not always respond to surveys, and previous attempts to collect data ran into response rates of as low as 6 percent, making their results unrepresentative. Enter the U.S. Census's Educational Quality of the Workforce National Employer Survey, first conducted in 1994, which collected data on business practices from a nationally representative sample of more than 1,500 workplaces.

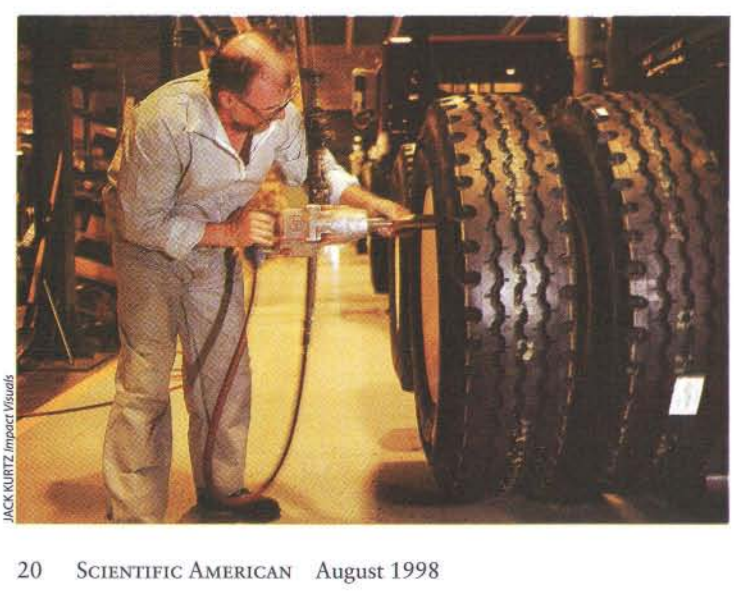
Lynch and Black correlated the survey data with other statistics that detailed the productivity of each business in the sample. They took as their "typical" establishment a nonunion company with limited profit sharing and without TQM or other formal quality-enhancing methods. (Unionized firms constituted about 20 percent of the sample, consistent with the waning reach of organized labor in the U.S.)

The average unionized establishment recorded productivity levels 16 percent higher than the baseline firm, whereas average nonunion ones scored 11 percent lower. One reason: most of the union shops had adopted so-called formal quality programs, in which up to half the workers meet regularly to discuss workplace issues. Moreover, production workers at these establishments shared in the firms' profits, and more than a quarter did their jobs in self-managed teams. Productivity in such union shops was 20 percent above baseline. That small minority of unionized workplaces still following the adversarial line recorded productivity 15 percent lower than the baseline—even worse than the nonunion average.

Are these productivity gains the result of high-performance management techniques rather than unionization? No, Lynch and Black say. Adoption of the same methods in nonunion establishments yielded only a 10 percent improvement in productivity over the baseline. The doubled gains in well-run union shops, Lynch contends, may result from the greater stake unionized workers have in their place of employment: they can accept or even propose large changes in job practice without worrying that they are cutting their own throats in doing so. (Lynch tells the opposing story of a high-tech company that paid its janitors a small bonus for suggesting a simple measure to speed nightly office cleaning—and then laid off a third of them.)

Even if a union cannot guarantee job security, she says, it enables workers to negotiate on a more or less equal footing. Especially in manufacturing, Lynch notes, unionized workplaces tend to have lower turnover. Consequently, they also reap more benefit from company-specific on-the-job training.

These documented productivity gains cast a different light on the declining percentage of unionized workers throughout the U.S. Are employers acting against their own interests when they



JACK KURTZ *Impact Visuals*

JOHNNY JOHNSON

**Compared with a nonunionized company with no quality management program.*

SOURCE: Lisa M. Lynch and Sandra E. Black

UNIONIZED EMPLOYEES are the most efficient workers if they are part of a "high-productivity" system, which includes quality management programs, profit sharing and regular discussions of workplace issues.

work to block unionization? Lynch believes that a follow-up survey, with initial analyses due out this winter, may help answer that question and others. Economists will be able to see how many of the previously sampled firms that have traditional management-labor relations managed to stay in business and to what extent the "corporate reengineering" mania of the past few years has paid off. Most serious reengineering efforts—the ones that aren't just downsizing by another name—lead to increased worker involvement, Lynch argues, if only because they require finding out how people actually do their jobs. Armed with that knowledge—and with the willing cooperation of their employees—firms may yet be able to break out of the productivity doldrums. *—Paul Wallich*

CYBER VIEW

Access Denied

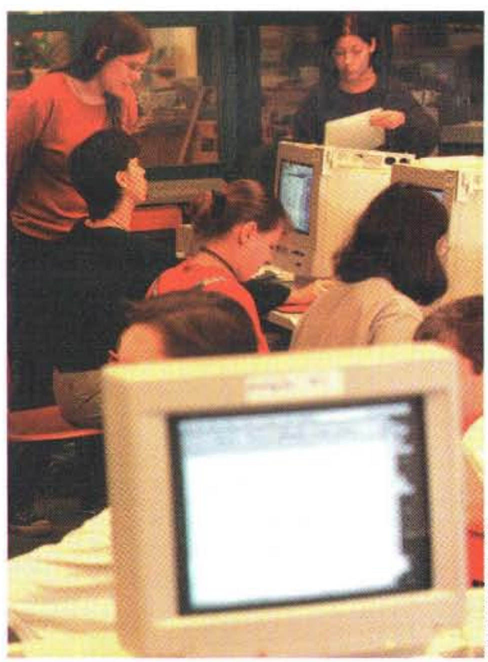
With all the advances toward equality for women, you've got to assume there are lots of women programmers, right? After all, it's not as though computer science is a discipline that requires more brawn than brains. But no: according to statistics compiled by Tracy K. Camp, then an assistant professor of computer science at the University of Alabama, the number of undergraduate degrees in computer science that is awarded to women has been shrinking steadily, both in real numbers and as a percentage of degrees awarded.

For instance, in the academic year 1980–81, women obtained 32.5 percent of the bachelor's degrees in computer science; the figure for 1993–94 was 28.4 percent, a drop of 12.6 percent. Calculated from the peak year in 1983–84, when 37.1 percent of the degrees went to women (representing 32,172 B.A. and B.S. degrees), the total decline is an alarming 23.5 percent.

This decline, Camp notes, is true even though other science and engineering fields boosted their recruitment of women. From 1980–81 to 1993–94, the percentage of physics degrees earned by women rose 36.6 percent (from 24.6 to 33.6 percent of recipients) and those in engineering by 44.7 percent (from 10.3 to 14.9 percent).

As in many other disciplines, there has been a "shrinking-pipeline" effect for women in computer science, who make up half of the high school computer science classes but who in 1993–94 constituted only 5.7 percent of full professorships in Ph.D.-granting departments. The percentage of M.S. degrees (logically) peaked in 1985–86, two years after the B.A. and B.S. degrees peaked; women earned 29.9 percent of the master's degrees awarded that year. Ph.D. degrees peaked at 15.4 percent in 1988–89 and again in 1993–94. Those figures decline further through the ranks of faculty members.

But Camp's research highlights a different problem: women are progressively staying away from computer science as a profession. Even though the absolute numbers of M.S. and Ph.D. degrees awarded in computer science have continued to increase, these numbers are likely to drop soon, reflecting the fewer undergraduate degrees awarded.



DICK BLUME *Image Works*

***WOMEN MAKE UP HALF THE STUDENTS** in high school computer science classes but often avoid the field later in life.*

The pattern is not unique to the U.S. In Britain, says Rachel Burnett, an information technology (IT) lawyer and partner with Masons Solicitors in London, the decline of women in computer science is even more marked. "The intake of women into university IT courses has declined," she says. "It used to be a third—it's now about 5 percent."

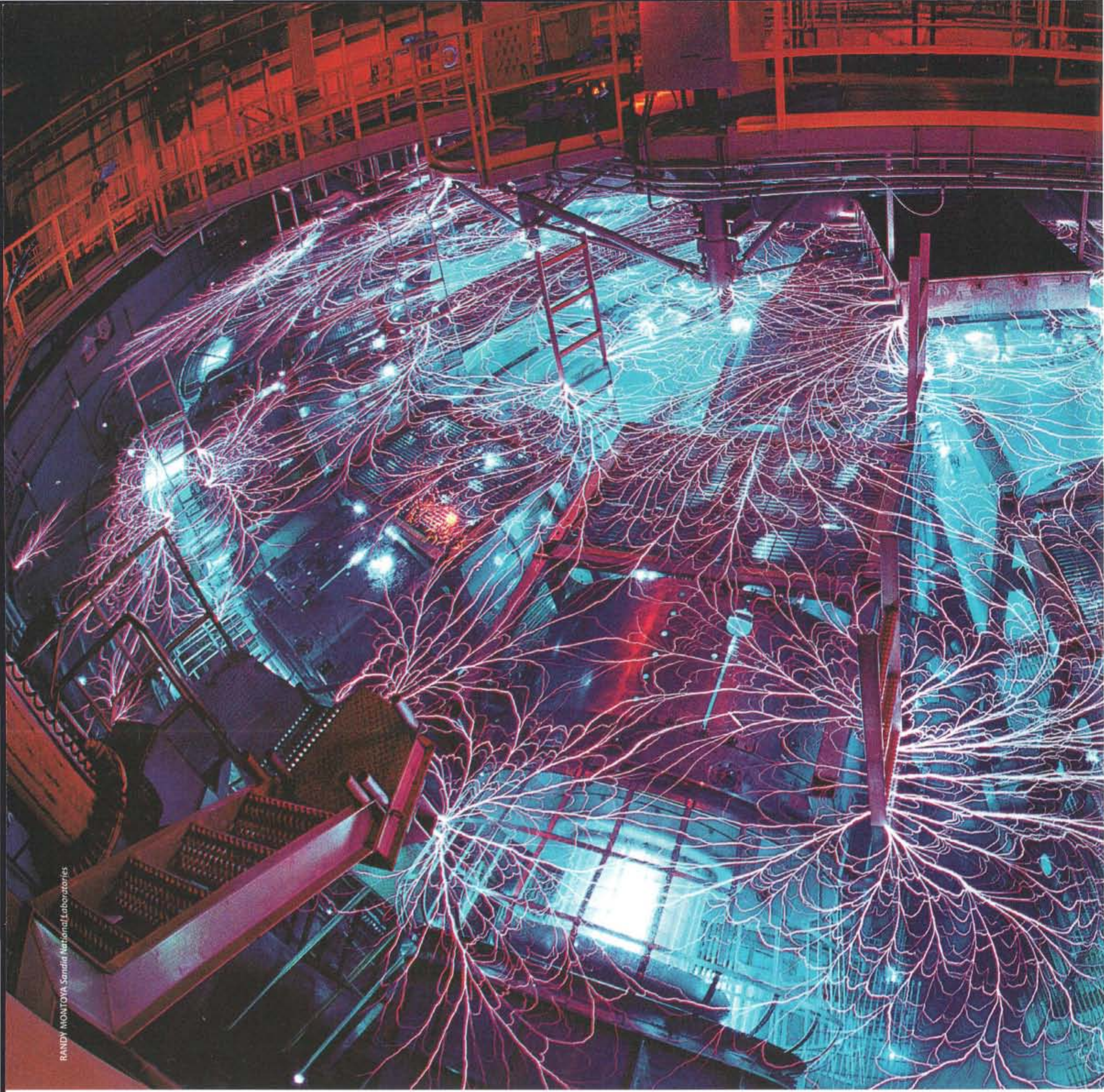
The question is why. To find out, Camp, who works on the Committee on Women in Computing for the Association for Computing Machinery (ACM), conducted a survey of ACM members. This is not, as she herself said when she presented the survey results at the ACM Policy '98 conference in Washington, D.C., this past May, exactly the right sample. To be realistic, you need to ask the people who didn't join the profession.

Still, the ACM respondents provided similar answers that plausibly explain the drop in women. They agreed that in the 1980s, computer games, which tend to be male-oriented, gave boys more computer experience. The nonstop long hours typical of programming jobs, the gender discrimination, the lack of role models and the antisocial image of the hacker have also tended to steer women to other areas. One issue that was not on the survey but that was brought up frequently by many conference participants was sexual harassment, which is perceived to be higher in the computer industry than in other fields.

Camp argues that this shrinking pipeline does matter. There is currently such a severe shortage of computer scientists that the industry recently lobbied Congress to grant more visas to would-be immigrants. And the quality and variety of software could only benefit from having a less homogeneous group of people creating it. Proposed solutions to raise the percentage of women who enter the profession include making more visible the female role models that exist, improving mentoring and encouraging equal access to computers for girls from kindergarten through the 12th grade. Perhaps then the existing pool of talent won't inadvertently seep away.

—Wendy M. Grossman in Washington, D.C.

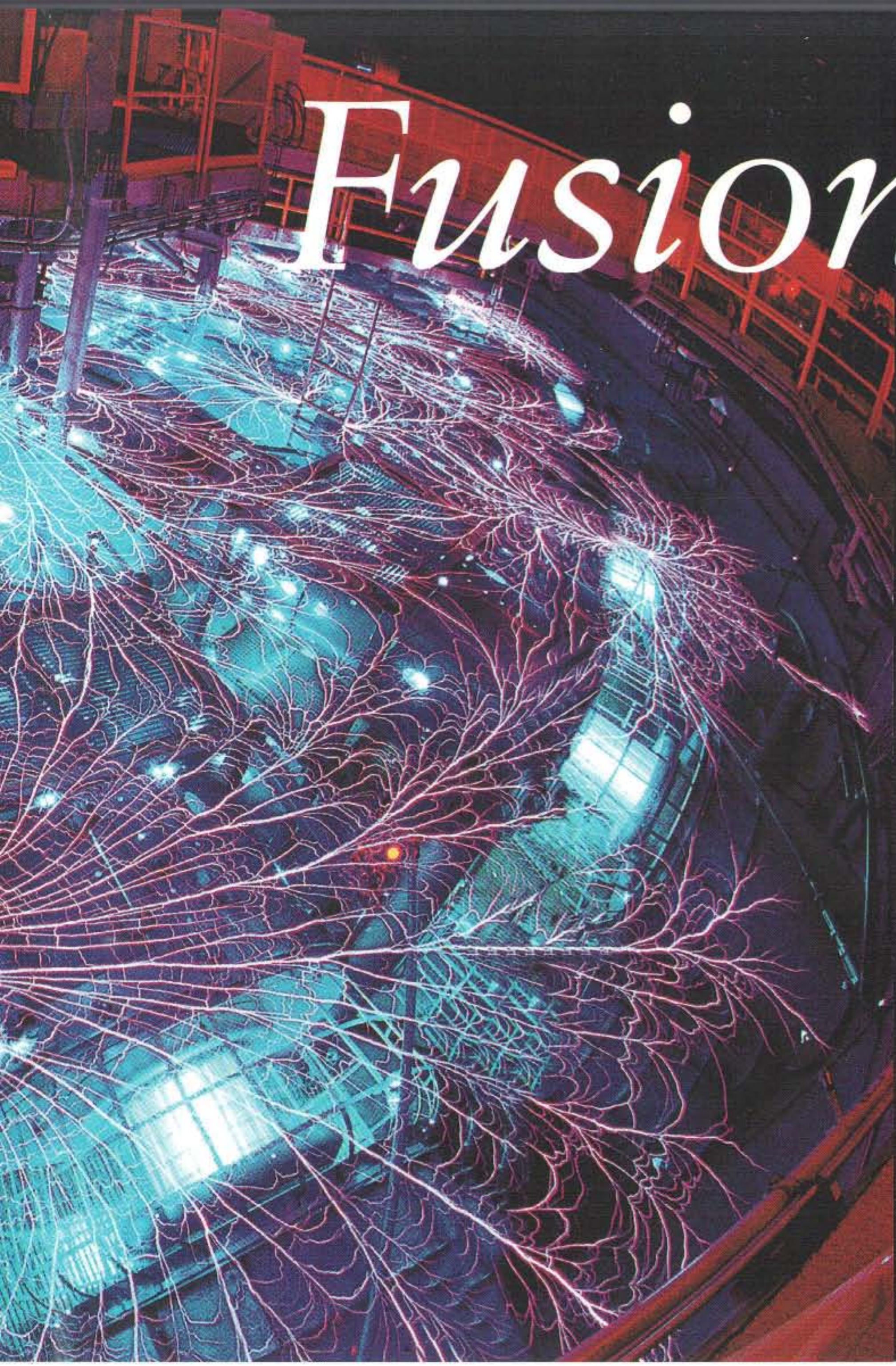
WENDY M. GROSSMAN, a freelance writer based in London, is the author of net.wars.



FIRING OF Z MACHINE produces a spectacular display. Z's primary transmission lines are submerged in water for insulation. A very small percentage of the huge energy used in firing the device escapes to the surface of the water in the form of large electrical discharges. As in a lightning strike, the high voltages break down the air-water interface, creating the visual discharges. The event lasts only microseconds. An automatic camera, with its shutter open, records the rapid discharges as a filigree of brilliant tracks.

Some things never change—or do they? In 1978 fusion research had been under way almost 30 years, and ignition had been achieved only in the hydrogen bomb. Nevertheless, I declared in *Scientific American* at the time that a proof of principle of laboratory fusion was less than 10 years away and that, with this accomplished, we could move on to fusion power plants [see "Fusion Power with Particle Beams," SCIENTIFIC AMERICAN, November 1978]. Our motivation, then as now, was the knowledge that a thimbleful of liquid heavy-hydrogen fuel could produce as much energy as 20 tons of coal.

Today researchers have been pursuing the Holy Grail of fusion for almost 50 years. Ignition, they say, is still "10 years away." The 1970s energy crisis is long forgotten, and the patience of our supporters is strained, to say the least. Less than three years ago I thought about pulling the plug on work at Sandia National Laboratories that was still a factor of 50 away from the power required to



Fusion and the Z Pinch

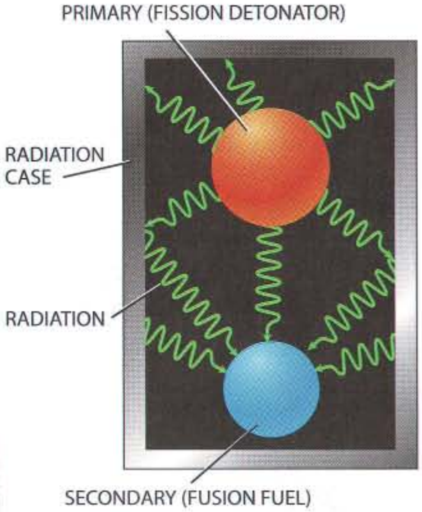
A device called the Z machine has led to a new way of triggering controlled fusion with intense nanosecond bursts of x-rays

by Gerold Yonas

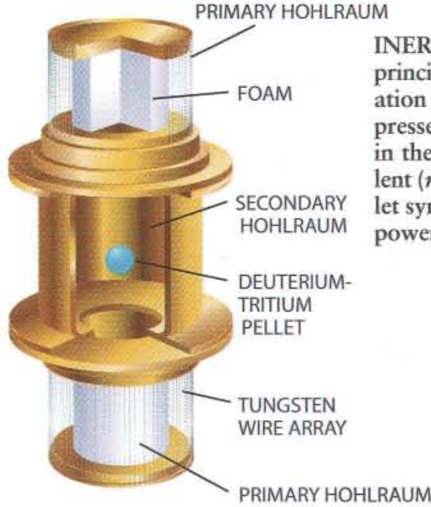
light the fusion fire. Since then, however, our success in generating powerful x-ray pulses using a new kind of device called the Z machine has restored my belief that triggering fusion in the laboratory may indeed be feasible in 10 years.

The hydrogen bomb provides the proof that fusion can be made to happen. In an H-bomb, radiation from an atomic fission explosion acts as a trigger, heating and compressing a fuel container to ignite and burn the hydrogen inside. That sounds simple, but causing a fusion reaction to ignite and burn means forcing together the nuclei of two forms of hydrogen, deuterium and tritium so that they fuse to form helium nuclei, giving off enormous energy. The compression must be done with almost perfect symmetry so that the hydrogen is squeezed uniformly to high density.

In the early decades of fusion research, the prospect of making a laboratory version of the H-bomb seemed remote. Instead efforts to control fusion relied on the principle of magnetic confinement, in which a powerful magnetic field traps a hot deuterium-tritium plasma long enough for fusion to begin. In 1991 deuterium-tritium fusion was achieved in this way by the Joint European Torus and later by Princeton University's Tokamak Fusion Test Reactor (TFTR). The next step on this road is the International Thermonuclear Experimental Reactor (ITER), a project involving the U.S., Europe, Japan and Russia. ITER's anticipated expense and technical difficulty, however, as well as disagreement



HYDROGEN BOMB



PROPOSED FUSION REACTION CHAMBER

MICHAEL GOODMAN

INERTIAL CONFINEMENT FUSION uses the principles of the hydrogen bomb (*left*), in which radiation from a fission bomb (called the primary) compresses and heats the fusion fuel, which is contained in the secondary. The minuscule laboratory equivalent (*right*) aims to bathe the peppercorn-size fuel pellet symmetrically in radiation and to concentrate the power into the pellet so that it implodes uniformly.

over where it should be built, have already slowed progress on the engineering design phase.

As far back as the early 1970s, researchers at Los Alamos, Lawrence Livermore and Sandia national laboratories turned their attention to another way of achieving fusion. In the inertial confinement approach, typified by the H-bomb, the idea is to use radiation to compress a pellet of hydrogen fuel. Whereas the H-bomb relies on radiation from an atomic bomb, the first attempts at laboratory inertial fusion made use of intense laser or electron beams to implode a fuel pellet.

The power then thought necessary to achieve ignition was much smaller than we now know it to be. By 1978 (as discussed in my earlier *Scientific American* article), the estimated requirement had risen to one million joules delivered in 10 nanoseconds to the outside of a peppercorn-size fuel pellet—a power demand of 100 terawatts and the equivalent of condensing several hours' worth of electricity use by half a dozen homes into a fraction of a second. To meet this need, we at Sandia and scientists in the Soviet Union began research with a novel technology called pulsed power.

In a pulsed power system, electrical energy is stored in capacitors and then discharged as brief pulses, which are made briefer still to increase the power in each pulse and compressed in space to increase the power density. These bursts of electromagnetic energy are then converted into intense pulses of charged particles or used to drive other devices. Laser fusion systems, in contrast, start with much longer electrical pulses, which are amplified and shaped within the lasing system itself. Pulsed power was seen as an attractive alternative to lasers because of its proved efficiency and low cost.

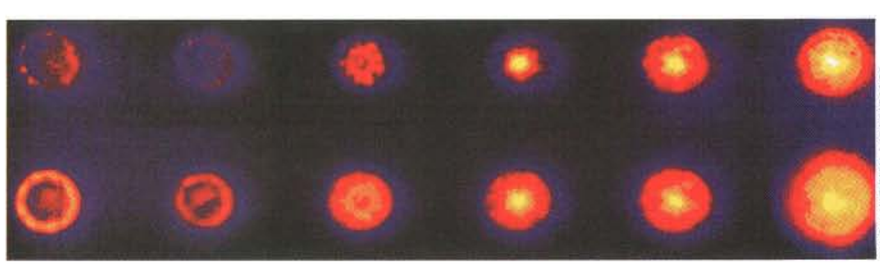
The technology began in 1964 at the U.K. Atomic Energy Authority and developed through the mid-1960s in the Soviet Union, the U.K. and the U.S. with support from the Energy and Defense departments. But the technique had limited power output, making it a dark horse in the race to fusion. Instead its preferred purpose was to simulate, in the laboratory, the effects of radiation on weapon components.

In 1973 funded programs in inertial confinement fusion began in the U.S. at Sandia under my direction and at other national laboratories and in the U.S.S.R. at the Kurchatov Institute under Leonid Rudakov. Since then, we have learned a tremendous amount about the technology for creating the power to reach ignition and the ignition requirements themselves, whether with lasers or with pulsed power. Decades of carefully diagnosed experiments using powerful lasers have improved and validated the computer codes that design fuel pellets. Today these simulations indicate that almost 500 terawatts and two million joules of radiation at a temperature of three million degrees for four nanoseconds are required to ignite the fuel.

Lasers can do this. After 13 years of research using the 30-kilojoule Nova laser, Lawrence Livermore is now building a much more powerful laser as the heart of the National Ignition Facility (NIF). If successful, the NIF will produce at least as much energy from fusion as the laser delivers to the pellet, but that will still not come close to producing the several 100-fold greater energy required to power the laser itself. That goal requires high yield—that is, fusion energy output much greater than the energy put into the laser. The NIF will take the next step toward high yield, but present laser technology is too expensive to go further.

Reviving the Z-Pinch

Just a few years ago, despite 25 years of effort, we at Sandia were still far from achieving fusion through pulsed power technology. The decision to continue the program was not easy, but our perseverance has recently been rewarded. For one, power output has grown enormously: pulsed power was producing one thousandth of a terawatt of radiation in the mid-1960s; recently we have reached 290 terawatts in experiments on the Z machine. We are confident that we can achieve high-yield fusion with radiation pulses of 1,000 terawatts, and we have come within a factor of three of that goal.



SANDIA NATIONAL LABORATORIES

X-RAYS are generated when a plasma from many fine tungsten wires collapses onto a carbon-deuterium straw placed on the axis of a Z-pinch. End-on views run from left to right at intervals of three nanoseconds. The top series shows x-rays that have energy greater than 800 electron volts; the bottom series shows x-rays around 200 electron volts in energy.

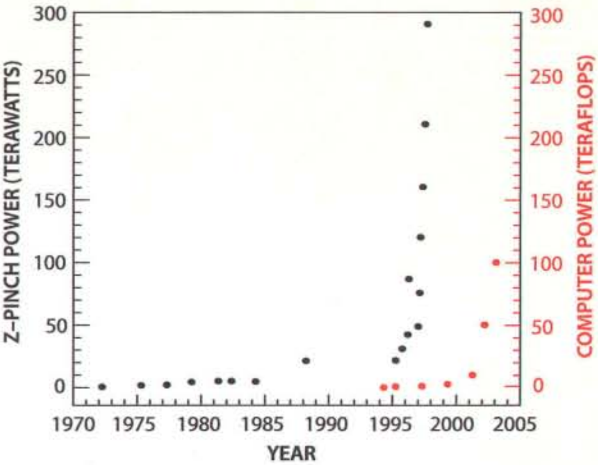
Equally important has been progress in concentrating the intense energy onto a tiny fuel pellet. In the 1970s we began with electron beams and, in the 1980s, switched to beams of ions, which should heat a target to higher temperatures. But charged particles are hard to steer and to focus tightly into beams. X-rays have, in principle, a much more promising characteristic: they can uniformly fill the space around a fuel container, just as heat in an oven envelops a turkey. The prospect of initiating fusion by using pulsed power systems to create intense bursts of x-rays within a small reaction chamber has now emerged from research on a concept called the Z-pinch, which dates back to the beginnings of magnetic confinement fusion research in the 1950s.

Originally, the Z-pinch was an attempt to initiate fusion by passing a strong electric current through deuterium gas. The current both ionizes the gas and generates a magnetic field that "pinches" the resulting plasma to high temperature and density along the current path, conventionally labeled the *z* axis. But the technique proved unable to compress a plasma uniformly: fluid instabilities break the plasma into blobs, making the Z-pinch inadequate to support fusion. The compression of the plasma, however, also generates x-rays, with energies up to 1,000 electron volts. For 30 years, research on Z-pinches in the U.S., the U.K. and the U.S.S.R. focused on optimizing the subkiloelectron-volt x-ray output, using those x-rays to test the response of materials and electronics to radiation from nuclear weapons.

The Z-pinch has now acquired new life as a way of initiating inertial fusion. In the past three years we have shown that by combining the efficiency and low cost of fast pulsed power with the simplicity and efficiency of the Z-pinch as a radiation source, we should reach ignition by using subkiloelectron-volt x-rays to compress a fusion fuel pellet. Moreover, the affordability of a pulsed power x-ray source should allow us to go beyond that, to efficient burn-up of the fuel and high yield.

To trigger fusion, the Z-pinch must be enclosed in a radiation chamber (or *hohlraum,* German for "cavity" or "hollow") that traps the x-rays. In one system we have explored, the Z-pinch would be placed in a primary hohlraum, with the fuel contained in a smaller, secondary hohlraum. In another method, the pellet would sit in low-density plastic foam at the center of the imploding pinch inside the primary hohlraum. The key is that the x-rays generated as the pinch crashes onto itself, either onto the z axis or onto the foam, are contained by the hohlraum so that they uniformly bathe the fuel pellet, just as the casing of an H-bomb traps the radiation from the atomic trigger. Experiments over the past three years show that both methods should work, because we can now make a Z-pinch that remains uniform and intact long enough to do the job.

POWER from wire-array Z-pinches increased slowly from 1970 to 1995. The rapid advances of the past two years in particular are the result of evolution to complex multiple-wire experiments using high-current accelerators and of improved computer modeling of plasmas. Red dots and scale show the increasing computer power available for simulations.



MICHAEL GOODMAN

Finding the Secret

What is different now compared with the long, earlier period of slow progress? Like Thomas Edison, who tried a thousand materials before finding the secret to the lightbulb, we discovered that the instability of the Z-pinch can be greatly reduced by extracting energy quickly, in the form of a short burst of x-rays, before instabilities destroy the geometry. In effect, the energy of the pinch is removed before it transforms into rapid and small-scale motions of plasma.

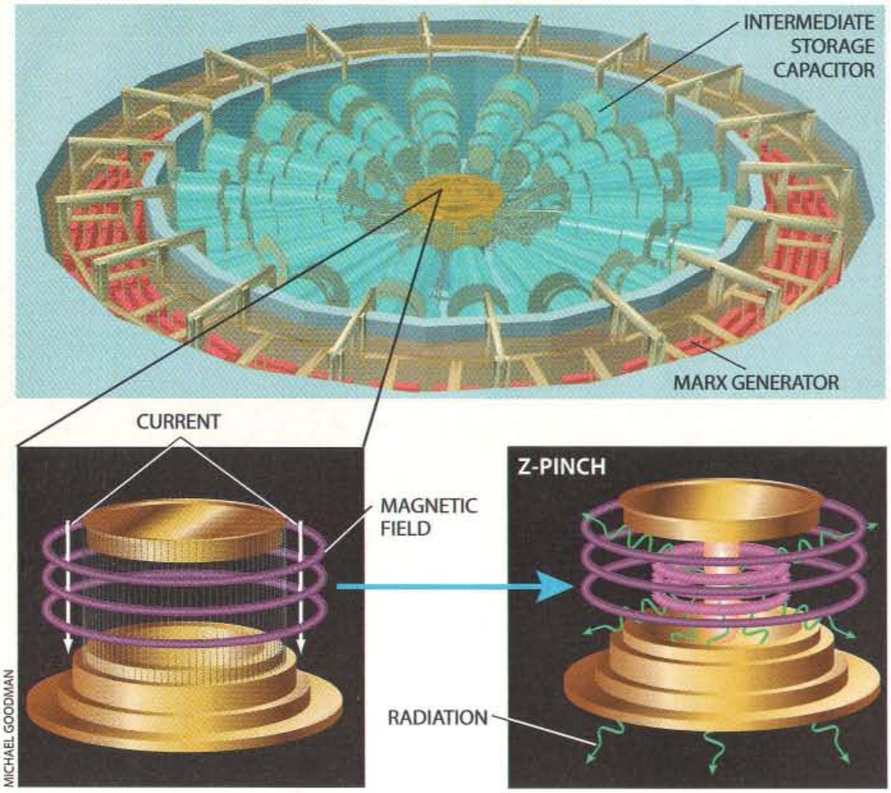
The instability that afflicts the pinch is the same one that makes a layer of vinegar poured carefully on top of less dense salad oil drop irregularly to the bottom of the jar. The instability is beneficial for salads when the jar is shaken, but it was a barrier to achieving fusion. From computer simulations by Darrell Peterson of Los Alamos and Melissa R. Douglas of Sandia, however, we knew that the more uniform the initial plasma, the more uniform and regular the pinch would be when it stagnated on axis and produced x-rays.

Researchers had tried many methods to make the plasma more uniform, such as using thin metal shells or hollow puffs of gas to conduct the electric current, but none met with great success. The breakthrough came at Sandia in 1995, when Thomas W. L. Sanford, using many fine aluminum wires, and then Christopher Deeney and Rick B. Spielman, using up to 400 fine tungsten wires, achieved the needed uniformity. Wire-array Z-pinches were first devised in the late 1970s at Physics International—a private company interested in generating x-rays as a source for testing radiation effects in the laboratory—for enhancing the energy output of one to five kiloelectron-volt x-rays. But the low-current accelerators then available could not deliver enough electric power to implode large numbers of many small wires.

After the 1995 experiments Barry M. Marder of Sandia suggested that the key was to have a number of wires arranged so that as they explode with the passage of the current, they merge to create a nearly uniform, imploding cylindrical plasma shell. Subsequent experiments at Sandia have shown that the desired hot central core is produced after the entire shell implodes onto a foam cylinder positioned on the *z* axis. Also, experiments at Cornell University indicate that each wire may not turn completely to plasma early on, as Marder's simulations suggest. Instead a cold core of wire may remain, surrounded by plasma, allowing the current flow to continue for a time and increasing the efficiency of the pinch.

These breakthroughs began three years ago at Sandia on the 10-million-ampere Saturn accelerator and, since October 1996, have continued on the 20-million-ampere Z machine, which now produces the world's most powerful and energetic x-ray pulses. In a typical experiment, we generate nearly two million joules of x-rays in a few nanoseconds,

PULSED-POWER GENERATOR



SEQUENTIAL CONCENTRATION of power in a pulsed-power facility begins in a circular array of 36 Marx generators, in which 90,000 volts charge a 5,000-cubic-meter capacitor bank in two minutes. Electrical pulses from the 36 modules enter a water-insulated section of intermediate storage capacitors, where they are compressed to a duration of 100 nanoseconds. Passing through a laser-triggered gas switch that synchronizes the 36 pulses to within one nanosecond, the combined pulse travels to the wire array (*far left*) along four magnetically insulated transmission lines that minimize loss of energy. The Z-pinch (*left*) forms as thousands of amps of electric current travel through wires one tenth the diameter of an average human hair. The illustrations on the opposite page demonstrate how x-rays from the pinch would implode a fuel pellet and initiate fusion in a schematic hohlraum design. In the proposed X-1 accelerator, the culmination would be the implosion, in about 10 nanoseconds, of a peppercorn-size fuel pellet to the size of the period at the end of this sentence.

for a power of more than 200 terawatts.

In a series of experiments beginning in November 1997, we increased the x-ray power by 45 percent, to 290 terawatts, by using a double-nested array of wires. The current vaporizes the outer array, and the magnetic field pushes the vaporized material inward. The faster-moving parts strike the inner array and are slowed, allowing the slower parts to catch up and sweep material into the inner array. This geometry reduces the instabilities in the implosion, and when the vaporized materials collide at the z axis they create a shorter pulse of x-rays than a single array can. The nested array has produced a radiation temperature of 1.8 million degrees.

In other experiments on Z, led by Arthur Toor of Livermore, foam layers surrounding a beryllium tube within a single array provide a slower, more symmetrical implosion of the Z-pinch plasma, also resulting in increased hohlraum temperatures. It took 40 years to get 40 terawatts of x-ray power from a Z-pinch. Now, in the past three years, we have moved much closer to our final goal of 1,000 terawatts and 16 million joules of x-rays, which should produce the three-million-degree hohlraum temperatures required for high-yield fusion. We knew that pulsed power should be more efficient and less costly than the laser approach, and indeed our Z accelerator now produces a total x-ray energy output equal to 15 percent of its electrical energy input; for Lawrence Livermore's Nova laser the equivalent efficiency is 0.1 percent. Design improvements could increase that figure to 0.5 percent at the NIF, but the inherent inefficiency of the laser process prevents such devices from achieving any higher efficiencies.

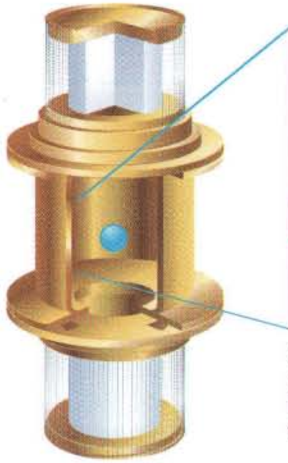
The Final Factor of Three

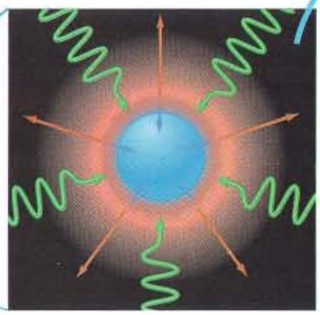
These intense x-ray pulses have many applications. The energies and powers attained with Z already allow laboratory measurements of material properties and studies of radiation transport at densities and temperatures that could previously be achieved only in underground nuclear explosions. These laboratory experiments, and the fusion yields that a higher-current device would permit, are part of the Department of Energy's stockpile stewardship program to guarantee the safety and reliability of aging nuclear weapons if the U.S. must use them in the future.

There are even astrophysical applications, because the x-ray sources powered by the Z machine produce plasmas similar to those in the outermost layers of a star. The light output from a type of pulsating star, the Cepheid variable, is now better understood because of data obtained by Paul T. Springer of Lawrence Livermore from Z-pinch experiments on the Saturn accelerator in 1996. We expect other data to lessen the mystery of stellar events such as supernovae. Laboratory plasmas also offer the potential for new studies in atomic physics and x-ray lasers.

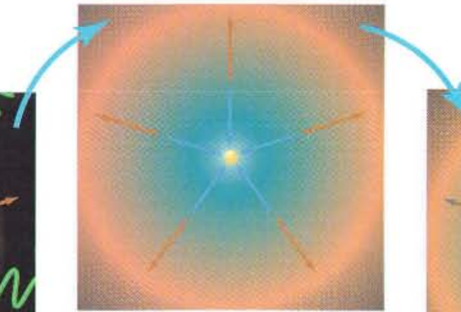
As an x-ray source, the Z-pinch is remarkably efficient and reproducible; repeated experiments yield the same magnitude of x-ray energy and power, even though we cannot predict in detail what happens. What we can forecast is scale: every time we double the current, the x-ray energy increases fourfold, following a simple square law. And as theoretically expected for thermal radiation, the pinch temperature increases as the square root of the current. If this physics holds true, another factor-of-three increase in current—to 60 million amperes—should allow us to achieve the energy, power and temperature needed to trigger fusion and reach high yield.

What questions must be resolved before that next step? The first is whether we can cram a factor-of-three-higher current into the same container. The reason that the enormous concentration of power into small cavities in a Z-pinch is possible at all was discovered almost 30 years ago at the Kurchatov Institute, at Physics International and at Sandia. Ordinarily, electric fields tend to disrupt

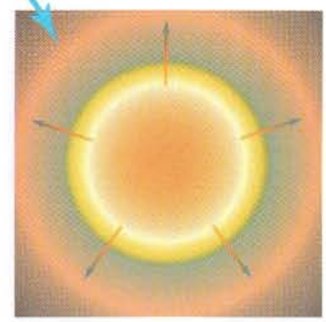




X-rays vaporize outer layer of fuel pellet, which bursts outward and imparts inward momentum to the hydrogen fuel.



Pellet implodes to $^{1}/_{1,000}$ to $^{1}/_{10,000}$ of its original volume, and fusion begins as temperature reaches 120 million degrees or more.



Wavefront of burning hydrogen expands outward and cools until fusion ceases.

MICHAEL GOODMAN

the current that generates them, but a phenomenon called magnetic insulation allows short, powerful electrical pulses to be transmitted along a channel between two metal surfaces without breaking down and with almost no energy loss. The magnetic field of the powerful pulse acts to contain the pulse itself, overcoming the electric field that would otherwise cause breakdown. In Z experiments by John L. Porter of Sandia in April 1998, a gap of 1.5 millimeters between a wire array and the surrounding stationary hohlraum wall remained open in the face of the intense radiation because of magnetic insulation, allowing the hohlraum temperature to reach 1.7 million degrees.

Fifty terawatts are transmitted into the tiny space between the wires and the hohlraum wall, and the power density reaches 25 terawatts per square centimeter. If we increase the power to 150 terawatts at 60 million amps, the power density would rise to 75 terawatts per square centimeter. That increase raises new questions, because the material pressure in the metal wall rises to 1.5 to three million atmospheres. Other questions are whether the efficiency of conversion to x-rays remains at the 15 percent level seen at 20 million amps, whether instabilities stay under control, and whether we can achieve the symmetry and shape of the radiation pulse onto the pellet that computer calculations suggest are needed.

Another important step is developing predictive models to scale the complex physics. The two-dimensional simulations available today have provided a great deal of insight into the physics of pinches but, even though restricted to two dimensions, require tremendous computer power. Simulation of the full three-dimensional magnetic, hydrodynamic and radiative character of the pinch is beyond our current capability, but advances in high-performance computing and diagnostics for x-ray imaging are rapidly catching up with our advances in radiated power. In 1998 we have been operating the Janus computer at 1.8 trillion floating-point (multiplication or division) operations per second (teraflops). Colleagues at Sandia and Los Alamos are developing a computer model to simulate the pinch physics and the transport of radiation to the pellet. With the advent of these tools on the new generation of supercomputers, we expect to continue our rapid progress in developing Z-pinches for fusion.

We at Sandia now hope to begin designing the next big step. At the end of March we requested approval from the Department of Energy to begin the conceptual design for the successor to Z, the X-1. This machine should give us 16 megajoules of radiation and, we believe, allow us to achieve high yield. Quoting a cost is premature, but we expect it will be in the neighborhood of $400 million. It is important to remember that Z, X-1 and the NIF are still research tools. Z may achieve fusion conditions; the NIF should achieve ignition; and X-1, building on the lessons of the NIF, should achieve high yield—but none of these experiments is expected to provide a commercial source of electrical power.

As Yogi Berra pointed out, "It is tough to make predictions, especially about the future," but if we can get started soon on design and construction of the next big step, we really think we can do the job in 10 years. SA

The Author

GEROLD YONAS is vice president of systems, science and technology at Sandia National Laboratories. His career began in 1962 at the Jet Propulsion Laboratory in Pasadena, Calif., while he was studying at the California Institute of Technology for his doctorate in engineering science and physics. Yonas joined Sandia in 1972. From 1984 to 1986 he provided technical management to the Strategic Defense Initiative Organization, serving as its first chief scientist. He was president of Titan Technologies in La Jolla, Calif., from 1986 to 1989, when he rejoined Sandia. He has published extensively in the fields of intense particle beams, inertial confinement fusion, strategic defense technologies and technology transfer.

Further Reading

NEUTRON PRODUCTION IN LINEAR DEUTERIUM PINCHES. O. A. Anderson, W. R. Baker, S. A. Colgate, J. Ise and R. V. Pyle in *Physical Review*, Vol. 110, No. 6, pages 1375–1387; 1958.

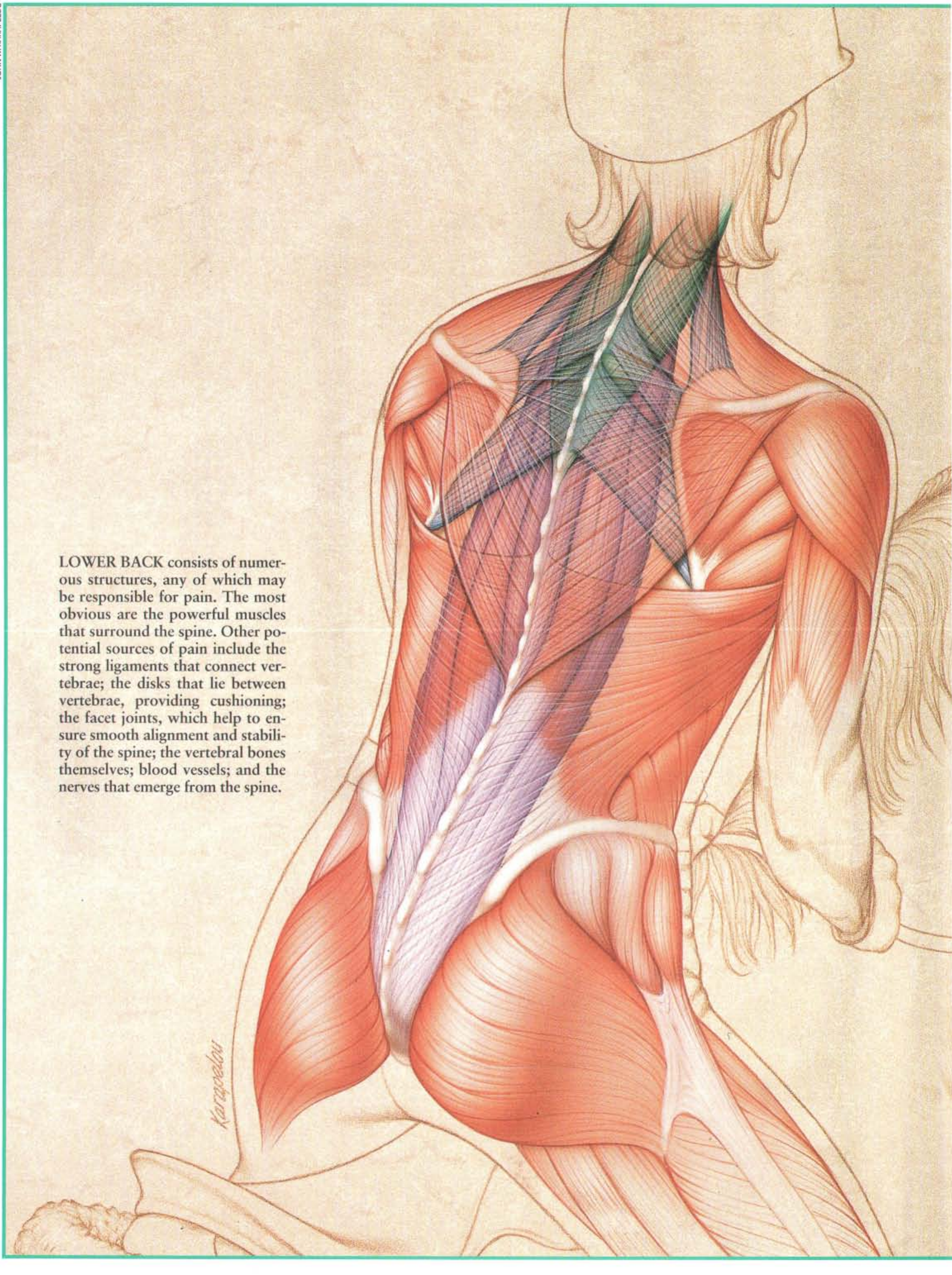
X-RAYS FROM Z-PINCHES ON RELATIVISTIC ELECTRON BEAM GENERATORS. N. R. Pereira and J. Davis in *Journal of Applied Physics*, Vol. 64, No. 3, pages R1–R27; 1988.

FAST LINERS FOR INERTIAL FUSION. V. P. Smirnov in *Plasma Physics and Controlled Fusion*, Vol. 33, No. 13, pages 1697–1714; November 1991.

DARK SUN: THE MAKING OF THE HYDROGEN BOMB. Richard Rhodes. Simon & Schuster, 1995.

Z-PINCHES AS INTENSE X-RAY SOURCES FOR HIGH-ENERGY DENSITY PHYSICS APPLICATIONS. M. Keith Matzen in *Physics of Plasmas*, Vol. 4, No. 5, Part 2, pages 1519–1527; May 1997.

JOHN W. KARAPELOU



LOWER BACK consists of numerous structures, any of which may be responsible for pain. The most obvious are the powerful muscles that surround the spine. Other potential sources of pain include the strong ligaments that connect vertebrae; the disks that lie between vertebrae, providing cushioning; the facet joints, which help to ensure smooth alignment and stability of the spine; the vertebral bones themselves; blood vessels; and the nerves that emerge from the spine.

Low-Back Pain

Low-back pain is at epidemic levels. Although its causes are still poorly understood, treatment choices have improved, with the body's own healing power often the most reliable remedy

by Richard A. Deyo

The catalogue of life's certainties is usually limited to death and taxes. A more realistic list would include low-back pain. Up to 80 percent of all adults will eventually experience back pain, and it is a leading reason for physician office visits, for hospitalization and surgery, and for work disability. The annual combined cost of back pain–related medical care and disability compensation may reach $50 billion in the U.S. Clearly, back pain is one of society's most significant nonlethal medical conditions. And yet the prevalence of back pain is perhaps matched in degree only by the lingering mystery accompanying it.

Consider the following paradox. The American economy is increasingly postindustrial, with less heavy labor, more automation and more robotics, and medicine has consistently improved diagnostic imaging of the spine and developed new forms of surgical and nonsurgical therapy. But work disability caused by back pain has steadily risen. Calling a physician a back-pain expert, therefore, is perhaps faint praise—medicine has at best a limited understanding of the condition. In fact, medicine's reliance on outdated ideas may have actually contributed to the problem. Old concepts were supported only by weak evidence such as physiological inferences and case reports, rather than by clinical findings from rigorous controlled trials.

The good news is that most back-pain patients will substantially and rapidly recover, even when their pain is severe. This prognosis holds true regardless of treatment method or even without treatment. Only a minority of patients with back pain will miss work because of it. Most patients who do leave work return within six weeks, and only a small percentage never return to their jobs. (At a given time, about 1 percent of the work force is chronically disabled because of back problems.) Overall, then, prospects for patients with acute back pain are quite good. The bad news is that recurrences are common; a majority of patients will experience them. Fortunately, these recurrences tend to play out much as the original incidents did, and most patients recover again quickly and spontaneously.

Sources of Pain

Low-back pain is a symptom that may signal various conditions affecting structures in the low back. Part of the mystery of back pain comes from the diagnostic challenge of determining its cause in a mechanical and biochemical system of multiple parts, all of which are subject to insult. Injuries to the muscles and ligaments may contribute, as may arthritis in the facet joints or disks. A herniated (or "slipped") disk, in which the soft inner cushioning material protrudes through the disk's outer rim and irritates an adjacent nerve root, can be the source of pain. Or the culprit might be spinal stenosis, a narrowing of the spinal canal that can cause a pinched nerve; stenosis usually accompanies aging and wear of the disks, the facet joints and the ligaments in the spinal canal.

Back pain also may be the result of congenital abnormalities of the spine. These odd structures are often asymptomatic but may cause trouble if severe enough. Diseases of other parts of the anatomy, such as the kidneys, pancreas, aorta or sex organs, can be responsible as well. Finally, back pain may be a

symptom of serious underlying diseases such as cancer, bone infections or rare forms of arthritis. Fortunately, such critical causes are extremely rare; about 98 percent of back-pain patients suffer from injury, usually temporary, to the muscles, ligaments, bones or disks.

The physical complexity of the lower back combines with another vexing reality to hinder diagnosis of the cause of pain: only a weak association exists between symptoms, imaging results, and anatomic or physiological changes. Under these circumstances, most diagnostic evaluations focus on excluding extreme causes of pain—such as cancer or infection—that can be more precisely identified or on determining whether a patient has a pinched or irritated nerve. Up to 85 percent of patients with low-back pain are then left without a definitive diagnosis, a nuts-and-bolts reason for their pain. Most patients cannot recall a specific incident that brought on their suffering, and heavy lifting or injuries, though risk factors, do not account for most episodes. Back pain often just seems to happen, and the medical community, reflecting this vagueness, has by no means reached a consensus as to the causes of garden-variety cases.

Some commonplace back pain is probably related to stress. A study published in May by Astrid Lampe and her colleagues at the University of Innsbruck revealed a connection between stressful life events and occurrences of back pain. Previous work by Lampe found that patients without a definite physical reason for low-back pain perceived life as more stressful than a control group of back-pain patients who had definite physical damage. John E. Sarno of the Rusk Institute of Rehabilitation Medicine at New York University Medical Center has concluded that unresolved emotionally charged states produce physical tension that in turn causes pain. In fact, he asserts that this variety of back pain actually serves to distract patients from the potential distress of confronting their psychological conflicts; Sarno has successfully treated selected patients with psychological counseling.

Simple muscle soreness from physical activity very likely causes some back pain, as does the natural wear and tear on disks and ligaments that creates microtraumas to those structures, especially with age. Determining the cause of a given individual's pain, however, often remains more art than science. With spontaneous recovery the rule—once serious disease is eliminated as a factor—pinpointing an exact cause may not even be necessary in most cases.

Diagnostic Challenges

The inadequacy of definitively diagnosing the cause of back pain led my colleague Daniel C. Cherkin of Group Health Cooperative of Puget Sound and my research group at the University of Washington to conduct a national survey of physicians from different specialties. We offered standardized patient descriptions and asked our subjects how they would manage these hypothetical patients. Reflecting the uncertainty in the state of the art, recommendations varied enormously. The results can be summed up by the subtitle of our publication of the survey results: "Who You See Is What You Get." For example, rheumatologists were twice as likely as physicians of other specialties to order laboratory tests in a search for arthritic conditions. Neurosurgeons were twice as likely to ask for imaging tests that would uncover herniated disks. And neurologists were three times more inclined to seek the results of electromyograms that would implicate the nerves. If patients are confused, they are not alone.

Until recently, doctors relied on spine x-rays, often performing one on every patient with low-back pain. Various studies have revealed multiple problems with this approach. First, a 10-year Swedish research effort demonstrated that at least for adults under age 50, x-rays added little of diagnostic value to office examinations, with unexpected findings in only about one of every 2,500 patients x-rayed.

Second, epidemiological research revealed that many conditions of the spine that often received blame for pain were actually unrelated to symptoms. Large numbers of pain-free people have been x-rayed in preemployment medical exams and for military induction in some countries, and multiple studies determined that many spine abnormalities were as common in asymptomatic people as in those with pain. X-rays can therefore be quite misleading.

Third, low-back x-rays unavoidably involve exposing sex organs to large doses of ionizing radiation, more than 1,000 times greater than that associated with a chest x-ray. Last, even highly experienced radiologists interpret the same x-rays differently, leading to uncertainty and even inappropriate treatment. The latest clinical guidelines for evaluating back pain thus recommend that x-rays be limited to specific patients, such as those who have suffered major injuries in a fall or automobile accident.

Medical experts hoped that improved diagnostic imaging instrumentation, such as computed tomographic (CT) scanning and magnetic resonance imaging (MRI), would make possible more precise diagnoses for most back-pain patients. This promise has been illusory. One important reason is that, as in the x-ray studies, alarming abnormalities are found in pain-free people.

A 1990 study by Scott D. Boden of the George Washington University Medical Center and his colleagues looked at 67 individuals who said they had never had any back pain or sciatica (leg pain from low-back conditions). Herniated disks often get cited as the reason for a patient's pain, but MRI found them in one fifth of pain-free study subjects under age 60. Half that group had a bulging disk, a less severe condition also often blamed for pain. Of adults older than 60, more than a third have a herniated disk visible with MRI, nearly 80 percent have a bulging disk and nearly everyone shows some age-related disk degeneration. Spinal stenosis, rare in younger adults, occurred in about one fifth of the over-60, pain-free group. A similar study of 98 pain-free people, published in 1994 by Michael N. Brant-Zawad-

Primary Reasons for Adult Visits to Office-Based Physicians

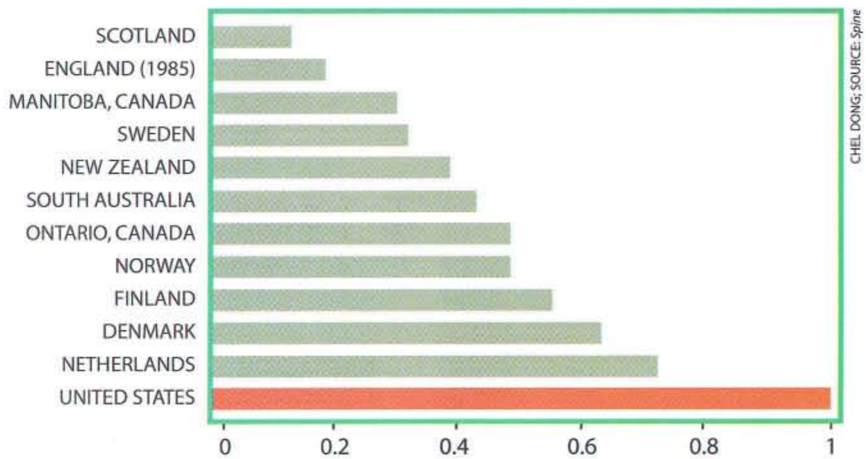
RANK

1. HYPERTENSION (HIGH BLOOD PRESSURE)
2. PREGNANCY CARE
3. CHECKUPS, WELL CARE
4. UPPER RESPIRATORY INFECTIONS (COLDS)
5. LOW-BACK PAIN
6. DEPRESSION AND ANXIETY
7. DIABETES

SOURCE: National Ambulatory Medical Care Survey, 1980–90

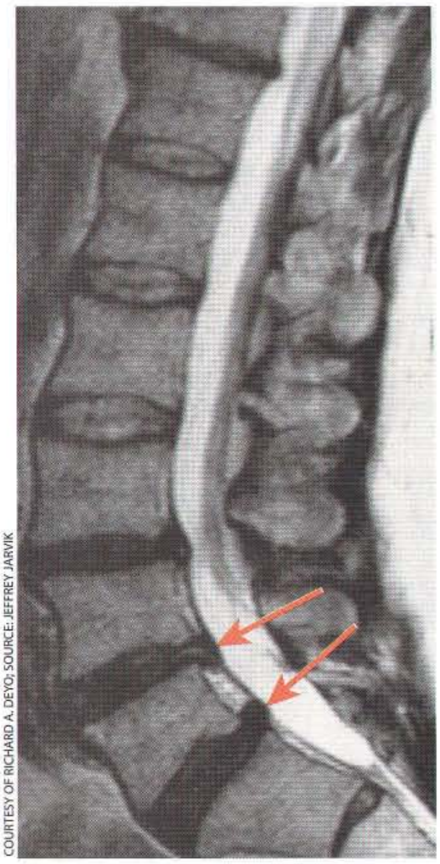
CHEL DONG

Ratio of Back-Surgery Rates in Selected Places, 1988–89



SURGERY RATES for back pain vary widely from country to country, with the U.S. having more than five times as many back operations performed as England. Varying cultural attitudes play a subtle role in the experience and the treatment of back pain, of which surgery is but an obvious example. In Oman, for example, back pain was common, but disability from back pain was rare until the introduction of Western medicine, according to a 1986 report to the minister of health.

zki of Hoag Memorial Hospital in Newport Beach, Calif., and his colleagues, revealed that about two thirds had abnormal disks. Detecting a herniated disk on an imaging test therefore proves only one thing conclusively: the patient has a herniated disk.



These findings suggest that many red herrings confuse imaging interpretation and that at least for some, spine abnormalities are purely coincidental and do not cause pain. Moreover, even the best imaging tests fail to identify the simple muscle spasm or injured ligament probably responsible for pain in a substantial percentage of back patients. All this imaging perplexity caused one orthopedic surgeon to remark, "A diagnosis based on MRI in the absence of objective clinical findings may not be the cause of a patient's pain, and an attempt at operative correction could be the first step toward disaster." In other words, the office examination is at least as important as the imaging test, and surgery for patients whose back pain is associated only with abnormal imaging results can be unnecessary if not downright detrimental. Many physicians now advocate CT scans and MRI only for those patients who are already surgical candidates for other reasons.

Complicating the situation still further is the fact that most patients with acute low-back pain simply get better—

ABNORMAL MRI SCAN from pain-free subject illustrates one of the great pitfalls in diagnosing a cause for low-back pain. Numerous studies have shown that large numbers of asymptomatic people also have disk bulges or herniations. This subject has two herniated disks (*arrows*) and additional disk abnormalities.

and quickly. A study comparing treatment outcomes found no differences in functional recovery times among patients who saw chiropractors, family doctors or orthopedic surgeons. Cost, on the other hand, varied substantially, with family doctors costing least and surgeons most. The Hippocratic admonition "First, do no harm" may be the most important counsel with regard to this condition—the favorable natural history of acute low-back pain is hard to beat.

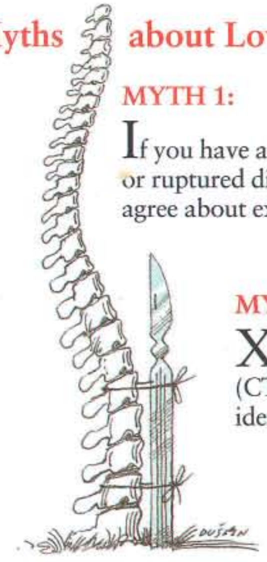
Extended bed rest was once regarded as the standard therapy. This approach was based on the rationale that some patients experience at least transient relief when lying down, as well as on the physiological observation that pressures in the intervertebral disks are lowest when patients are prone. But a guilty-looking disk may be innocent, and most patients improve naturally. Nevertheless, recommendations of one to two weeks of strict bed rest were the norm until about 10 years ago. Bed rest's fall from favor has been almost as dramatic as the reversal in status suffered by that former favorite of primary care, bloodletting. Extended bed rest is now considered anathema, and resuming normal activities as much as possible may be the best option for patients with acute back pain.

Watchful Waiting as Treatment

When bed rest was still the standard, my group tested it by comparing seven days of bed rest with just two days. The results were striking. After three weeks and three months, there were no differences in pain relief, in days of limited activity, in daily functioning or in satisfaction with care. The only difference was that, obviously, patients with longer bed rest missed more work. Severity of a patient's pain, duration of pain, and abnormalities found in the office examination offered no predictive value for how long the patient would be off the job. In fact, data analysis showed that the only factor that predicted the duration of the patient's absence from work was our recommendation for how long to stay in bed.

Other studies have confirmed and extended these findings. Four days of bed rest turns out to be no more effective than two days—or even no bed rest at all. The fear that activity would exacerbate the situation and delay recovery proved to be unfounded. In fact, pa-

Myths about Low-Back Pain



MYTH 1:

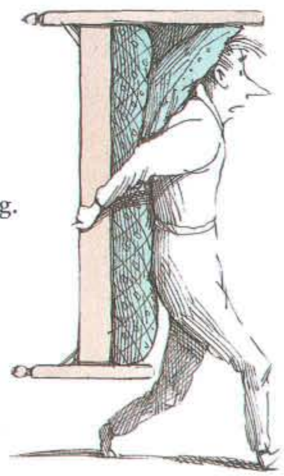
If you have a slipped disk (also known as a herniated or ruptured disk) you must have surgery. Surgeons agree about exactly who should have surgery.

MYTH 2:

X-ray and newer imaging tests (CT and MRI scans) can always identify the cause of pain.

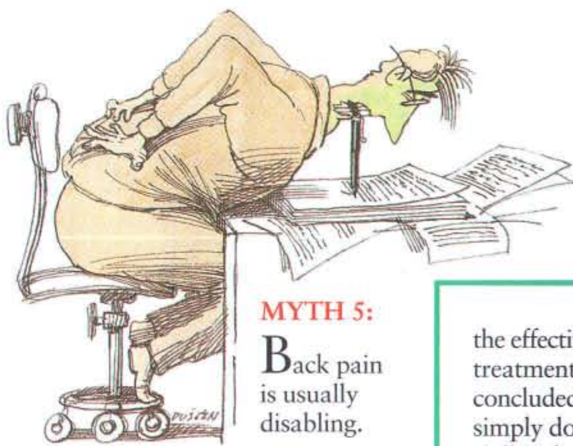
MYTH 3:

If your back hurts, you should take it easy until the pain goes away.



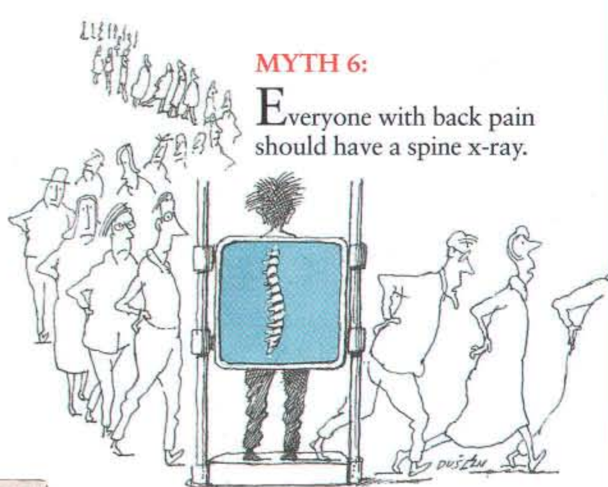
MYTH 4:

Most back pain is caused by injuries or heavy lifting.



MYTH 5:

Back pain is usually disabling.



MYTH 6:

Everyone with back pain should have a spine x-ray.

MYTH 7:

Bed rest is the mainstay of therapy.

ILLUSTRATIONS BY DUSAN PETRICIC

tients with acute back pain who continue routine activities as normally as possible do better than those who try either bed rest or immediate exercise. Studies have shown that people who remained active despite acute pain experienced less future chronic pain (defined as pain lasting three months or more) and used fewer health care services than patients who rested and waited for the pain to diminish. (The fact that bed rest is ineffective does not mean that everyone can return to their normal jobs immediately, however. Some people with physically demanding jobs may be unable to go back to their normal work as quickly as people with more sedentary occupations. Nevertheless, it is often useful to have patients with back pain return to some form of light work until they have recovered more fully.)

Recent research has also challenged the effectiveness of other types of passive treatment. For example, several studies concluded that traction for the low back simply does not work. More controversially, there is growing evidence that transcutaneous electrical nerve stimulation (TENS), which delivers mild electric current to the painful area, has little if any long-term benefit. Similarly, injections of the facet joints with cortisone-like drugs appear to be no more effective than injections with saline solution.

In contrast, there is growing evidence for exercise as an important part of the prevention and treatment of back problems for those suffering from either chronic or acute back pain. No single exercise is best, and effective programs combine aerobics for general fitness with specific training to improve the strength and endurance of the back muscles.

An exhaustive review of clinical studies of exercise and back pain found that structured exercise programs prevented recurrences and reduced work absences in patients with acute pain who regularly took part soon after an episode of back pain had subsided. The preventive power of exercise was stronger than the effect of education (for example, how to lift) or of abdominal belts that limit spine motion. Patients experiencing chronic pain also benefited from exercise. In contrast to acute back-pain sufferers, who did better during a pain episode by resuming normal activities than through exercise, chronic back-pain patients substantially improved by exercising even with their pain.

The inability of conventional medical practice to "cure" a large percentage of back-pain patients has no doubt led the condition to be a major reason patients seek various forms of alternative treatment, including chiropractic care and acupuncture. Chiropractic is the most common choice, and evidence accumulates that spinal manipulation may indeed be an effective short-term pain remedy for patients with recent back problems. Whether chiropractic or other alternative treatments can impart long-term pain relief remains unclear. The normal recovery from back pain most likely leads to a belief in whatever treatment is employed and probably accounts for the large number of therapeutic options with passionate advocates.

At the other end of the strategic spectrum is surgery. Most specialists agree that disk surgery is appropriate only when there is a combination of a definite disk hernia on an imaging test, a corresponding pain syndrome, signs of nerve root irritation and failure to respond to six weeks of nonsurgical treatment. For patients with these findings, surgery can offer faster pain relief. Unfortunately, patients who do not meet

all these standards also often go under the knife, and there is extensive literature on failed low-back surgery. Indeed, if the pain is not actually from disk herniation, surgical repair of a disk cannot be expected to end it.

Surgical Interventions

The scapegoating of the herniated disk deserves further reflection. Herniated disks are most common in adults between ages 30 and 50, and most patients whose pain is actually caused by a disk herniation have leg pain with numbness and tingling as the primary symptom; their back pain is often less severe. A positive MRI should only support a physical examination that investigates a constellation of effects—such as nerve root irritation, reflex abnormalities and limited sensation, muscle strength and leg mobility—to implicate the disk definitively as the factor in pain.

Recent studies show that even for patients with a herniated disk, spontaneous recovery is the rule. Studies using repeated MRI revealed that the herniated part of the disk often shrinks naturally over time, and about 90 percent of patients will experience gradual improvement over a period of six weeks. Thus, only about 10 percent of patients with a symptomatic disk herniation would appear to require surgery. And because most back pain is not caused by herniated disks, the actual proportion of all back-pain patients who are surgical candidates is only about 2 percent.

Herniated disks nonetheless remain the most common reason for back surgery. A long-term follow-up study of 280 patients, performed by Henrik Weber of Ullevaal Hospital in Oslo and published in 1983, raises serious questions about the enthusiasm for surgical intervention. Although patients who had surgery had faster pain relief than did patients treated conservatively, the differences evaporated over time. At the four- and 10-year follow-ups, the two groups of patients were virtually indistinguishable. Thus, reasonable people might have preferences for different medical interventions, and there is growing recognition that these preferences should be an important consideration in treatment decisions.

Spinal stenosis is the most common reason for back surgery in those over age 65. National hospital survey data show stenosis correction to be the most rapidly increasing form of back surgery. Surgery for herniated disks increased 39 percent between 1979 and 1990; stenosis surgeries increased 343 percent. Reasons for this rapid rise are unclear but may simply reflect the ability of the new CT and MRI scans to reveal stenosis. Unfortunately, the indications for surgery in this condition are even less clear-cut than they are for herniated disks. As a result, there are enormous variations, even within the U.S., in rates of surgery for spinal stenosis. For example, by analyzing Medicare claims, my group found approximately 30 stenosis surgeries in Rhode Island for every 100,000 people older than 65 but 132 in Utah.

Surgery for this condition is more complex than simple disk surgery. Spinal stenosis tends to occur at multiple levels within the spine rather than at a single level, as is usually true for herniated disks. Furthermore, these patients are older and therefore more susceptible to complications of surgery. In addition, we know less about the long-term effectiveness of surgical and nonsurgical approaches for treating spinal stenosis than we do about management of herniated disks. Because symptoms of spinal stenosis often remain stable for years at a time, decisions are rarely urgent, and the preferences of the patient should again play an important role.

Classifying as trivial a condition that annually drives millions of Americans to their knees and drains $50 billion from the economy would be a mistake. A collective shrug at the condition, however, may be the most appropriate, albeit unsatisfying, societal attitude. Nearly everyone will have back pain, and we should perhaps simply accept it as part of normal life. Once serious conditions get ruled out, a sufferer is usually best served by simply attempting to cope as well as possible with a condition that will almost certainly improve in days or a few weeks. The wide variability in surgical recommendations should make all back-pain experts circumspect, and the patient's wishes should carry considerable weight in treatment choice.

The mysterious nature and economic cost of back pain are driving a growing interest in research, and the coming years may reveal the fundamental aspects of this problem in more detail. In the meantime, for most back-pain patients the stereotypical physician advisory to "take two aspirin and call me in the morning" comes to mind. A richer and better course of action might be to take pain relievers as needed, stay in good overall physical condition, keep active through an acute attack if at all possible and monitor the condition for changes over a few days or a week. Back pain's power to inflict misery is great, but that power is usually transient. In most cases, time and perseverance will carry a patient through to recovery. SA

The Author

RICHARD A. DEYO is a general internist with a long-standing interest in clinical research on low-back problems and is a professor in the departments of medicine and of health services at the University of Washington. He received his M.P.H. from its School of Public Health and Community Medicine in 1981 and his M.D. from the Pennsylvania State University School of Medicine in 1975. Deyo has studied various therapies for low-back pain, including bed rest, exercise regimens and transcutaneous electrical nerve stimulation (TENS).

Further Reading

A New Clinical Model for the Treatment of Low-Back Pain. Gordon Waddell in *Spine,* Vol. 12, No. 7, pages 632–644; 1987.

Cost, Controversy, Crisis: Low Back Pain and the Health of the Public. Richard A. Deyo, Daniel Cherkin, Douglas Conrad and Ernest Volinn in *Annual Review of Public Health,* Vol. 12, pages 141–156; 1991.

Physician Variation in Diagnostic Testing for Low Back Pain. Daniel C. Cherkin, Richard A. Deyo, Kimberly Wheeler and Marcia A. Ciol in *Arthritis & Rheumatism,* Vol. 37, No. 1, pages 15–22; January 1994.

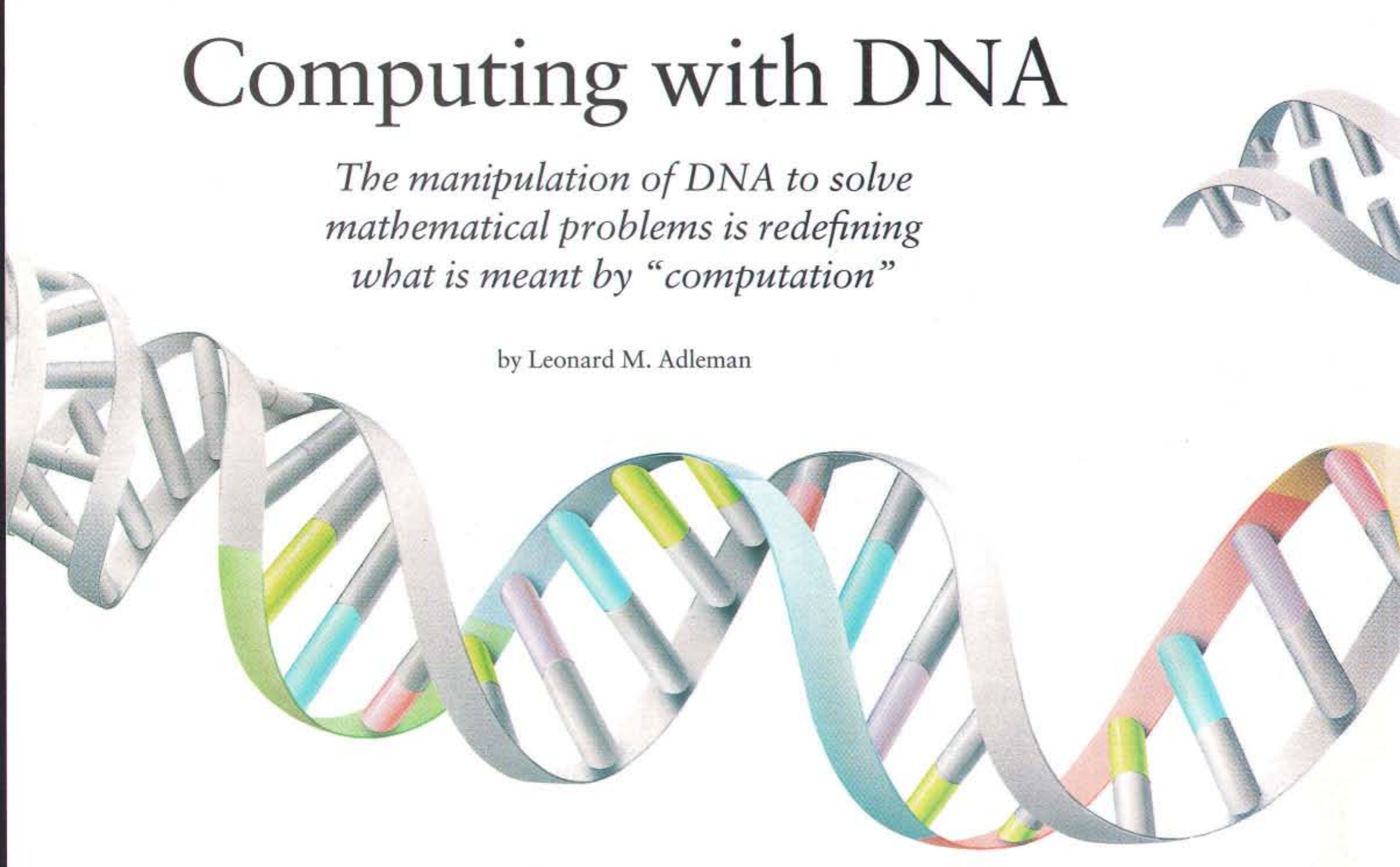
Magnetic Resonance Imaging of the Lumbar Spine in People without Back Pain. Maureen C. Jensen, Michael N. Brant-Zawadzki, Nancy Obuchowski, Michael T. Modic, Dennis Malkasian and Jeffrey S. Ross in *New England Journal of Medicine,* Vol. 331, No. 2, pages 69–73; July 14, 1994.

The Mindbody Prescription: Healing the Body, Healing the Pain. John E. Sarno. Warner Books, 1998.

Computing with DNA

The manipulation of DNA to solve mathematical problems is redefining what is meant by "computation"

by Leonard M. Adleman



Computer. The word conjures up images of keyboards and monitors. Terms like "ROM," "RAM," "gigabyte" and "megahertz" come to mind. We have grown accustomed to the idea that computation takes place using electronic components on a silicon substrate.

But must it be this way? The computer that you are using to read these words bears little resemblance to a PC. Perhaps our view of computation is too limited. What if computers were ubiquitous and could be found in many forms? Could a liquid computer exist in which interacting molecules perform computations? The answer is yes. This is the story of the DNA computer.

Rediscovering Biology

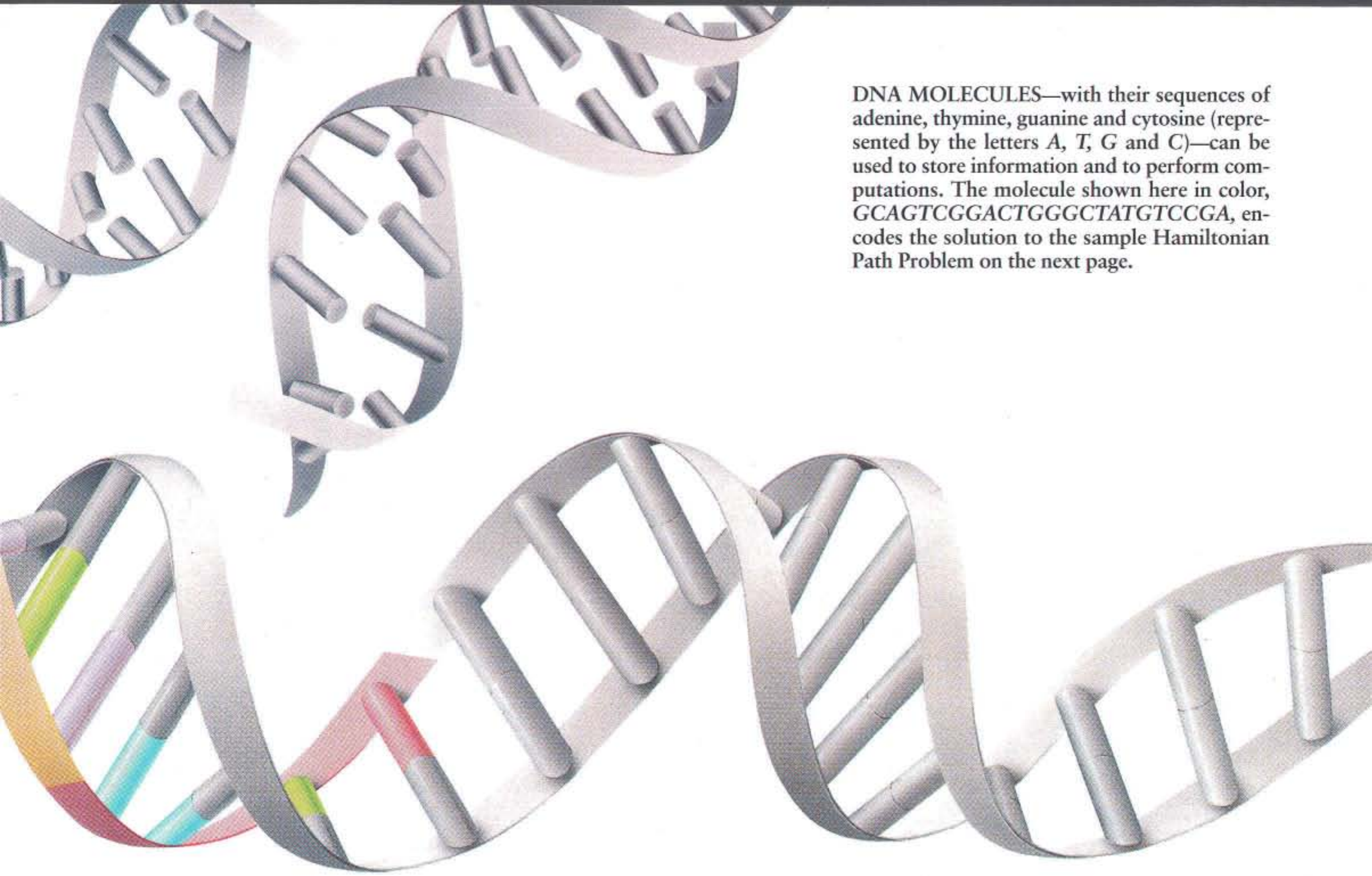
My involvement in this story began in 1993, when I walked into a molecular biology lab for the first time. Although I am a mathematician and computer scientist, I had done a bit of AIDS research, which I believed and still believe to be of importance [see "Balanced Immunity," by John Rennie; SCIENTIFIC AMERICAN, May 1993]. Unfortunately, I had been remarkably unsuccessful in communicating my ideas to the AIDS research community. So, in an effort to become a more persuasive advocate, I decided to acquire a deeper understanding of the biology of HIV. Hence, the molecular biology lab. There, under the guidance of Nickolas Chelyapov (now chief scientist in my own laboratory), I began to learn the methods of modern biology.

I was fascinated. With my own hands, I was creating DNA that did not exist in nature. And I was introducing it into bacteria, where it acted as a blueprint for producing proteins that would change the very nature of the organism.

During this period of intense learning, I began reading the classic text *The Molecular Biology of the Gene,* co-authored by James D. Watson of Watson-Crick fame. My concept of biology was being dramatically transformed. Biology was no longer the science of things that smelled funny in refrigerators (my view from undergraduate days in the 1960s at the University of California at Berkeley). The field was undergoing a revolution and was rapidly acquiring the depth and power previously associated exclusively with the physical sciences. Biology was now the study of information stored in DNA—strings of four letters: *A, T, G* and *C* for the bases adenine, thymine, guanine and cytosine—and of the transformations that information undergoes in the cell. There was mathematics here!

Late one evening, while lying in bed reading Watson's text, I came to a description of DNA polymerase. This is the king of enzymes—the maker of life. Under appropriate conditions, given a strand of DNA, DNA polymerase produces a second "Watson-Crick" complementary strand, in which every *C* is replaced by a *G,* every *G* by a *C,* every *A* by a *T* and every *T* by an *A.* For example, given a molecule with the sequence *CATGTC,* DNA polymerase will produce a new molecule with the sequence *GTACAG.* The polymerase enables DNA to reproduce, which in turn allows cells to reproduce and ultimately allows you to reproduce. For a strict reductionist, the replication of DNA by DNA polymerase is what life is all about.

DNA polymerase is an amazing little nanomachine, a single molecule that



DNA MOLECULES—with their sequences of adenine, thymine, guanine and cytosine (represented by the letters *A, T, G* and *C*)—can be used to store information and to perform computations. The molecule shown here in color, *GCAGTCGGACTGGGCTATGTCCGA,* encodes the solution to the sample Hamiltonian Path Problem on the next page.

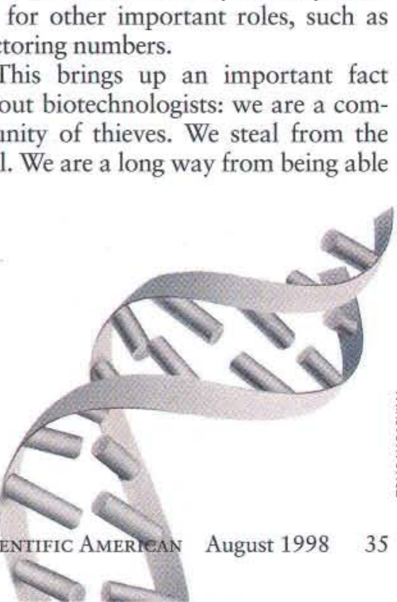
"hops" onto a strand of DNA and slides along it, "reading" each base it passes and "writing" its complement onto a new, growing DNA strand. While lying there admiring this amazing enzyme, I was struck by its similarity to something described in 1936 by Alan M. Turing, the famous British mathematician. Turing—and, independently, Kurt Gödel, Alonzo Church and S. C. Kleene—had begun a rigorous study of the notion of "computability." This purely theoretical work preceded the advent of actual computers by about a decade and led to some of the major mathematical results of the 20th century [see "Unsolved Problems in Arithmetic," by Howard DeLong; SCIENTIFIC AMERICAN, March 1971; and "Randomness in Arithmetic," by Gregory J. Chaitin; SCIENTIFIC AMERICAN, July 1988].

For his study, Turing had invented a "toy" computer, now referred to as a Turing machine. This device was not intended to be real but rather to be conceptual, suitable for mathematical investigation. For this purpose, it had to be extremely simple—and Turing succeeded brilliantly. One version of his machine consisted of a pair of tapes and a mechanism called a finite control, which moved along the "input" tape reading data while simultaneously moving along the "output" tape reading and writing other data. The finite control was programmable with very simple instructions, and one could easily write a program that would read a string of *A, T, C* and *G* on the input tape and write the Watson-Crick complementary string on the output tape. The similarities with DNA polymerase could hardly have been more obvious.

But there was one important piece of information that made this similarity truly striking: Turing's toy computer had turned out to be universal—simple as it was, it could be programmed to compute anything that was computable at all. (This notion is essentially the content of the well-known "Church's thesis.") In other words, one could program a Turing machine to produce Watson-Crick complementary strings, factor numbers, play chess and so on. This realization caused me to sit up in bed and remark to my wife, Lori, "Jeez, these things could compute." I did not sleep the rest of the night, trying to figure out a way to get DNA to solve problems.

My initial thinking was to make a DNA computer in the image of a Turing machine, with the finite control replaced by an enzyme. Remarkably, essentially the same idea had been suggested almost a decade earlier by Charles H. Bennet and Rolf Landauer of IBM [see "The Fundamental Physical Limits of Computation"; SCIENTIFIC AMERICAN, July 1985]. Unfortunately, while an enzyme (DNA polymerase) was known that would make Watson-Crick complements, it seemed unlikely that any existed for other important roles, such as factoring numbers.

This brings up an important fact about biotechnologists: we are a community of thieves. We steal from the cell. We are a long way from being able

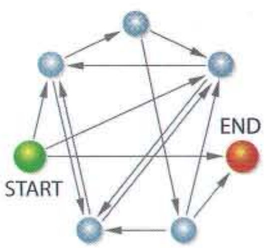


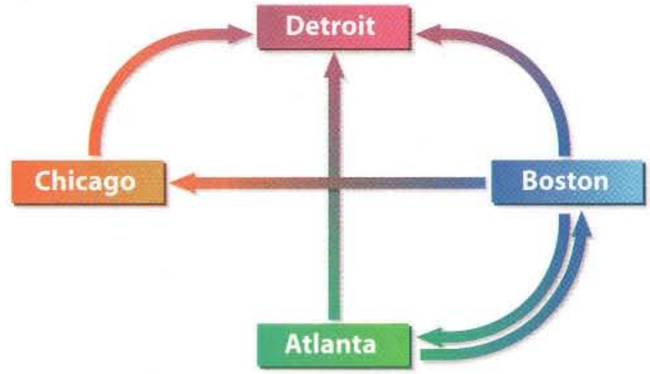
TOMO NARASHIMA

Hamiltonian Path Problem

Consider a map of cities connected by certain nonstop flights (*top right*). For instance, in the example shown here, it is possible to travel directly from Boston to Detroit but not vice versa. The goal is to determine whether a path exists that will commence at the start city (Atlanta), finish at the end city (Detroit) and pass through each of the remaining cities exactly once. In DNA computation, each city is assigned a DNA sequence (*ACTTGCAG* for Atlanta) that can be thought of as a first name (*ACTT*) followed by a last name (*GCAG*). DNA flight numbers can then be defined by concatenating the last name of the city of origin with the first name of the city of destination (*bottom right*).

The complementary DNA city names are the Watson-Crick complements of the DNA city names in which every *C* is replaced by a *G*, every *G* by a *C*, every *A* by a *T*, and every *T* by an *A*. (To simplify the discussion here, details of the 3′ versus 5′ ends of the DNA molecules have been omitted.) For this particular problem, only one Hamiltonian path exists, and it passes through Atlanta, Boston, Chicago and Detroit in that order. In the computation, this path is represented by *GCAGTCGGACTGGGCTATGTCCGA*, a DNA sequence of length 24. Shown at the left is the map with seven cities and 14 nonstop flights used in the actual experiment. —*L.M.A.*





CITY	DNA NAME	COMPLEMENT
ATLANTA	ACTTGCAG	TGAACGTC
BOSTON	TCGGACTG	AGCCTGAC
CHICAGO	GGCTATGT	CCGATACA
DETROIT	CCGAGCAA	GGCTCGTT

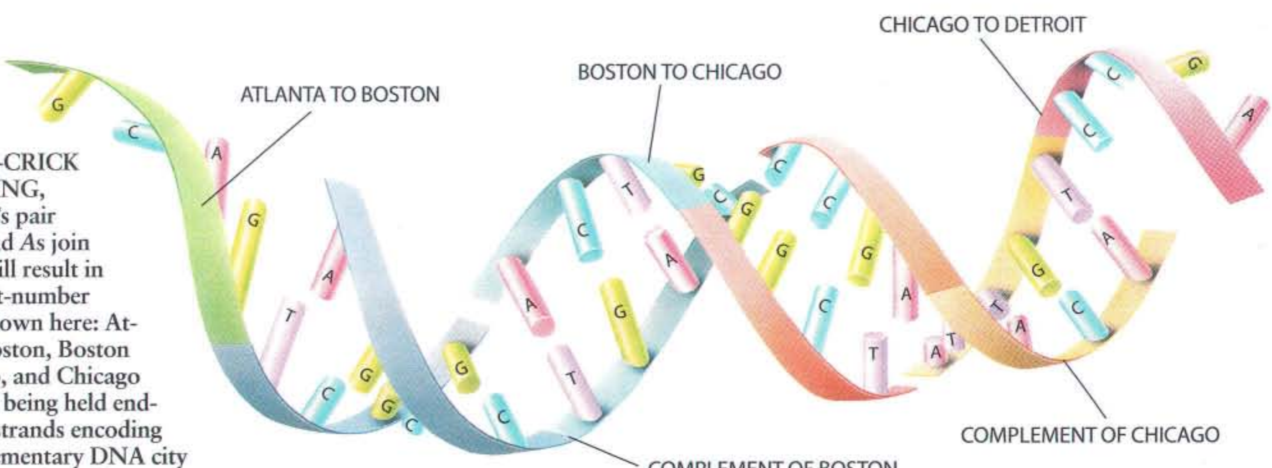
FLIGHT	DNA FLIGHT NUMBER
ATLANTA - BOSTON	GCAGTCGG
ATLANTA - DETROIT	GCAGCCGA
BOSTON - CHICAGO	ACTGGGCT
BOSTON - DETROIT	ACTGCCGA
BOSTON - ATLANTA	ACTGACTT
CHICAGO - DETROIT	ATGTCCGA

SLIM FILMS

to create de novo miraculous molecular machines such as DNA polymerase. Fortunately, three or four billion years of evolution have resulted in cells that are full of wondrous little machines. It is these machines, stolen from the cell, that make modern biotechnology possible. But a molecular machine that would play chess has apparently never evolved. So if I were to build a DNA computer that could do something interesting, I would have to do it with the tools that were at hand. These tools were essentially the following:

1. **Watson-Crick pairing.** As stated earlier, every strand of DNA has its Watson-Crick complement. As it happens, if a molecule of DNA in solution meets its Watson-Crick complement, then the two strands will anneal—that is, twist around each other to form the famous double helix. The strands are not covalently bound but are held together by weak forces such as hydrogen bonds. If a molecule of DNA in solution meets a DNA molecule to which it is not complementary (and has no long stretches of complementarity), then the two molecules will not anneal.

2. **Polymerases.** Polymerases copy information from one molecule into another. For example, DNA polymerase will make a Watson-Crick complementary DNA strand from a DNA template. In fact, DNA polymerase needs a "start



WATSON-CRICK ANNEALING, in which *C*s pair with *G*s and *A*s join with *T*s, will result in DNA flight-number strands (shown here: Atlanta to Boston, Boston to Chicago, and Chicago to Detroit) being held end-to-end by strands encoding the complementary DNA city names (shown here: Boston and Chicago).

signal" to tell it where to begin making the complementary copy. This signal is provided by a primer—a (possibly short) piece of DNA that is annealed to the template by Watson-Crick complementarity. Wherever such a primer-template pair is found, DNA polymerase will begin adding bases to the primer to create a complementary copy of the template.

3. **Ligases.** Ligases bind molecules together. For example, DNA ligase will take two strands of DNA in proximity and covalently bond them into a single strand. DNA ligase is used by the cell to repair breaks in DNA strands that occur, for instance, after skin cells are exposed to ultraviolet light.

4. **Nucleases.** Nucleases cut nucleic acids. For example, restriction endonucleases will "search" a strand of DNA for a predetermined sequence of bases and, when found, will cut the molecule into two pieces. EcoRI (from *Escherichia coli*) is a restriction enzyme that will cut DNA after the *G* in the sequence *GAATTC*—it will almost never cut a strand of DNA anywhere else. It has been suggested that restriction enzymes evolved to protect bacteria from viruses (yes, even bacteria have viruses!). For example, *E. coli* has a means (methylation) of protecting its own DNA from EcoRI, but an invading virus with the deadly *GAATTC* sequence will be cut to pieces. My DNA computer did not use restriction enzymes, but they have been used in subsequent experiments by many other research groups.

5. **Gel electrophoresis.** This is not stolen from the cell. A solution of heterogeneous DNA molecules is placed in one end of a slab of gel, and a current is applied. The negatively charged DNA molecules move toward the anode, with shorter strands moving more quickly than longer ones. Hence, this process separates DNA by length. With special chemicals and ultraviolet light, it is possible to see bands in the gel where the DNA molecules of various lengths have come to rest.

6. **DNA synthesis.** It is now possible to write a DNA sequence on a piece of paper, send it to a commercial synthesis facility and in a few days receive a test tube containing approximately 10^{18} molecules of DNA, all (or at least most) of which have the described sequence. Currently sequences of length approximately 100 can be reliably handled in this manner. For a sequence of length 20, the cost is about $25. The molecules are delivered dry in a small tube and appear as a small, white, amorphous lump.

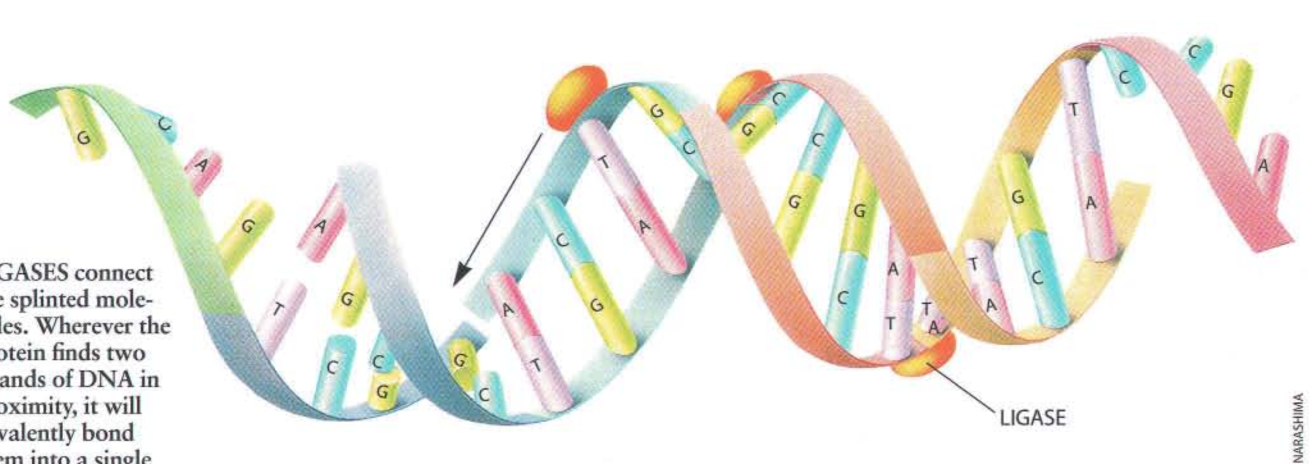
None of these appeared likely to help play chess, but there was another important fact that the great logicians of the 1930s taught us: computation is easy. To build a computer, only two things are really necessary—a method of storing information and a few simple operations for acting on that information. The Turing machine stores information as sequences of letters on tape and manipulates that information with the simple instructions in the finite control. An electronic computer stores information as sequences of zeros and ones in memory and manipulates that information with the operations available on the processor chip. The remarkable thing is that just about any method of storing information and any set of operations to act on that information are good enough.

Good enough for what? For universal computation—computing anything that can be computed. To get your computer to make Watson-Crick complements or to play chess, you need only start with the correct input information and apply the right sequence of operations—that is, run a program. DNA is a great way to store information. In fact, the cell has been using this method to store the "blueprint for life" for billions of years. Further, enzymes such as polymerases and ligases have been used to operate on this information. Was there enough to build a universal computer? Because of the lessons of the 1930s, I was sure that the answer was yes.

The Hamiltonian Path Problem

The next task was choosing a problem to solve. It should not appear to be contrived to fit the machine, and it should demonstrate the potential of this novel method of computation. The problem I chose was the Hamiltonian Path Problem.

William Rowan Hamilton was Astronomer Royal of Ireland in the mid-19th century. The problem that has come to bear his name is illustrated in the box on the opposite page. Let the arrows (directed edges) represent the nonstop flights between the cities (vertices) in the map (graph). For example, you can fly nonstop from Boston to Chicago but not from Chicago to Boston. Your job (the Hamiltonian Path Problem) is to determine if a sequence of connecting flights (a path) exists that starts in Atlanta (the start vertex) and ends in Detroit (the end vertex), while passing through each of the remaining cities (Boston and Chicago) exactly once. Such a path is called a Hamiltonian path. In the example shown on



LIGASES connect the splinted molecules. Wherever the protein finds two strands of DNA in proximity, it will covalently bond them into a single strand.

page 36, it is easy to see that a unique Hamiltonian path exists, and it passes through the cities in this order: Atlanta, Boston, Chicago, Detroit. If the start city were changed to Detroit and the end city to Atlanta, then clearly there would be no Hamiltonian path.

More generally, given a graph with directed edges and a specified start vertex and end vertex, one says there is a Hamiltonian path if and only if there is a path that starts at the start vertex, ends at the end vertex and passes though each remaining vertex exactly once. The Hamiltonian Path Problem is to decide for any given graph with specified start and end vertices whether a Hamiltonian path exists or not.

The Hamiltonian Path Problem has been extensively studied by computer scientists. No efficient (that is, fast) algorithm to solve it has ever emerged. In fact, it seems likely that even using the best currently available algorithms and computers, there are some graphs of fewer than 100 vertices for which determining whether a Hamiltonian path exists would require hundreds of years.

In the early 1970s the Hamiltonian Path Problem was shown to be "NP-complete." Without going into the theory of NP-completeness, suffice it to say that this finding convinced most theoretical computer scientists that no efficient algorithm for the problem is possible at all (though proving this remains the most important open problem in theoretical computer science, the so-called NP = P? problem [see "Turing Machines," by John E. Hopcroft; Scientific American, May 1984]). This is not to say that no algorithms exist for the Hamiltonian Path Problem, just no efficient ones. For example, consider the following algorithm:

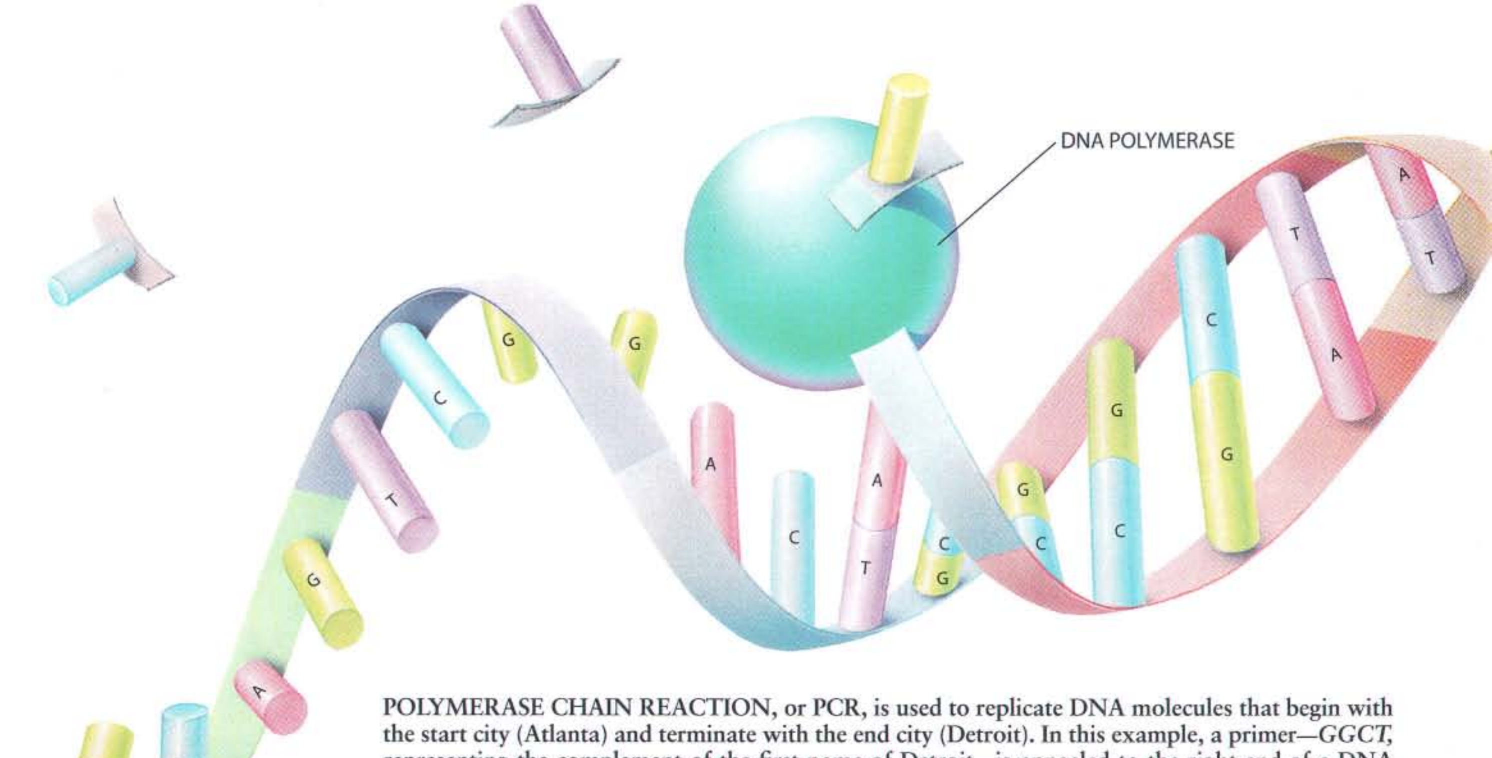
Given a graph with *n* vertices,

1. Generate a set of random paths through the graph.
2. For each path in the set:
   a. Check whether that path starts at the start vertex and ends with the end vertex. If not, remove that path from the set.
   b. Check if that path passes through exactly *n* vertices. If not, remove that path from the set.
   c. For each vertex, check if that path passes through that vertex. If not, remove that path from the set.
3. If the set is not empty, then report that there is a Hamiltonian path. If the set is empty, report that there is no Hamiltonian path.

This is not a perfect algorithm; nevertheless, if the generation of paths is random enough and the resulting set large enough, then there is a high probability that it will give the correct answer. It is this algorithm that I implemented in the first DNA computation.

Seven Days in a Lab

For my experiment, I sought a Hamiltonian Path Problem small enough to be solved readily in the lab, yet large enough to provide a clear "proof of principle" for DNA computation. I chose the seven-city, 14-flight map shown in the inset on page 36. A nonscientific study has shown that it takes about 54



POLYMERASE CHAIN REACTION, or PCR, is used to replicate DNA molecules that begin with the start city (Atlanta) and terminate with the end city (Detroit). In this example, a primer—*GGCT,* representing the complement of the first name of Detroit—is annealed to the right end of a DNA strand. The primer signals the polymerase to begin making a Watson-Crick complement of the strand. After the polymerase is done, the double helix is split into two strands so that Watson-Crick complements of each half can be made. This process is repeated to obtain a large number of copies of molecules that have the correct start and end cities. Gel electrophoresis is then used to isolate those molecules that have the right sequence length of 24.

seconds on average to find the unique Hamiltonian path in this graph. (You may begin now....)

To simplify the discussion here, consider the map on page 36, which contains just four cities—Atlanta, Boston, Chicago and Detroit—linked by six flights. The problem is to determine the existence of a Hamiltonian path starting in Atlanta and ending in Detroit.

I began by assigning a random DNA sequence to each city. In our example, Atlanta becomes *ACTTGCAG*, Boston *TCGGACTG* and so on. It was convenient to think of the first half of the DNA sequence as the first name of the city and the second half as the last name. So Atlanta's last name is *GCAG*, whereas Boston's first name is *TCGG*. Next, I gave each nonstop flight a DNA "flight number," obtained by concatenating the last name of the city of origin with the first name of the city of destination. In the example on page 36, the Atlanta-to-Boston flight number becomes *GCAGTCGG*.

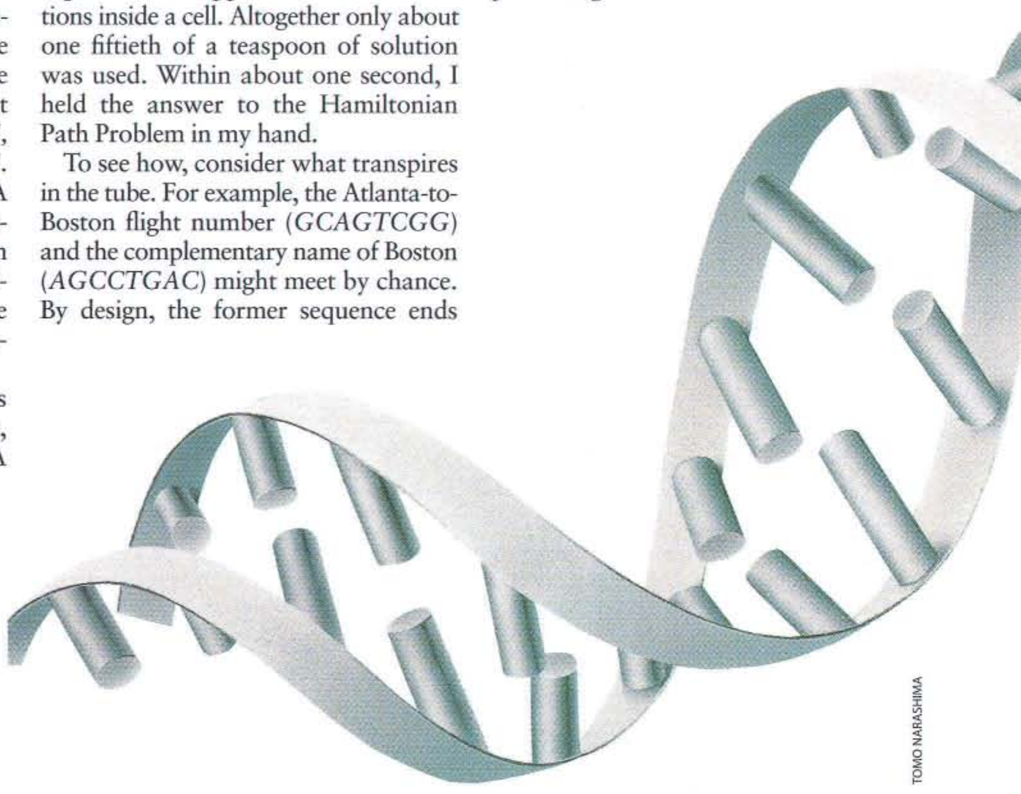
Recall that each strand of DNA has its Watson-Crick complement. Thus, each city has its complementary DNA name. Atlanta's complementary name becomes, for instance, *TGAACGTC*.

After working out these encodings, I had the complementary DNA city names and the DNA flight numbers synthesized. (As it turned out, the DNA city names themselves were largely unnecessary.) I took a pinch (about 10^{14} molecules) of each of the different sequences and put them into a common test tube. To begin the computation, I simply added water—plus ligase, salt and a few other ingredients to approximate the conditions inside a cell. Altogether only about one fiftieth of a teaspoon of solution was used. Within about one second, I held the answer to the Hamiltonian Path Problem in my hand.

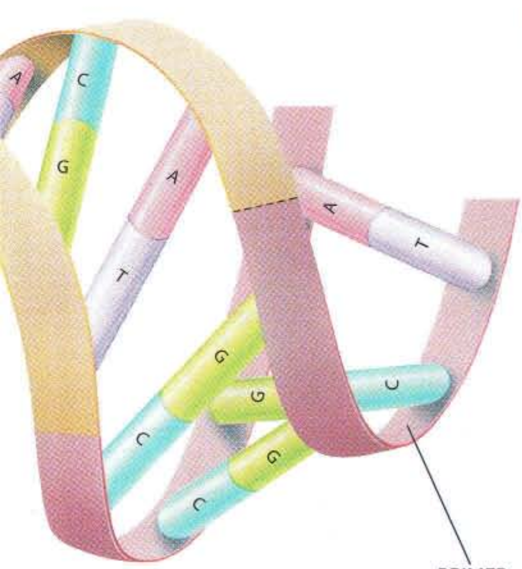
To see how, consider what transpires in the tube. For example, the Atlanta-to-Boston flight number (*GCAGTCGG*) and the complementary name of Boston (*AGCCTGAC*) might meet by chance. By design, the former sequence ends with *TCGG*, and the latter starts with *AGCC*. Because these sequences are complementary, they will stick together. If the resulting complex now encounters the Boston-to-Chicago flight number (*ACTGGGCT*), it, too, will join the complex because the end of the former (*TGAC*) is complementary to the beginning of the latter (*ACTG*). In this manner, complexes will grow in length, with DNA flight numbers splinted together by complementary DNA city names. The ligase in the mixture will then permanently concatenate the chains of DNA flight numbers. Hence, the test tube contains molecules that encode random paths through the different cities (as required in the first step of the algorithm).

Because I began with such a large number of DNA molecules and the problem contained just a handful of cities, there was a virtual certainty that at least one of the molecules formed would encode the Hamiltonian path. It was amazing to think that the solution to a mathematical problem could be stored in a single molecule!

Notice also that all the paths were created at once by the simultaneous interactions of literally hundreds of trillions of molecules. This biochemical reaction represents enormous parallel processing.

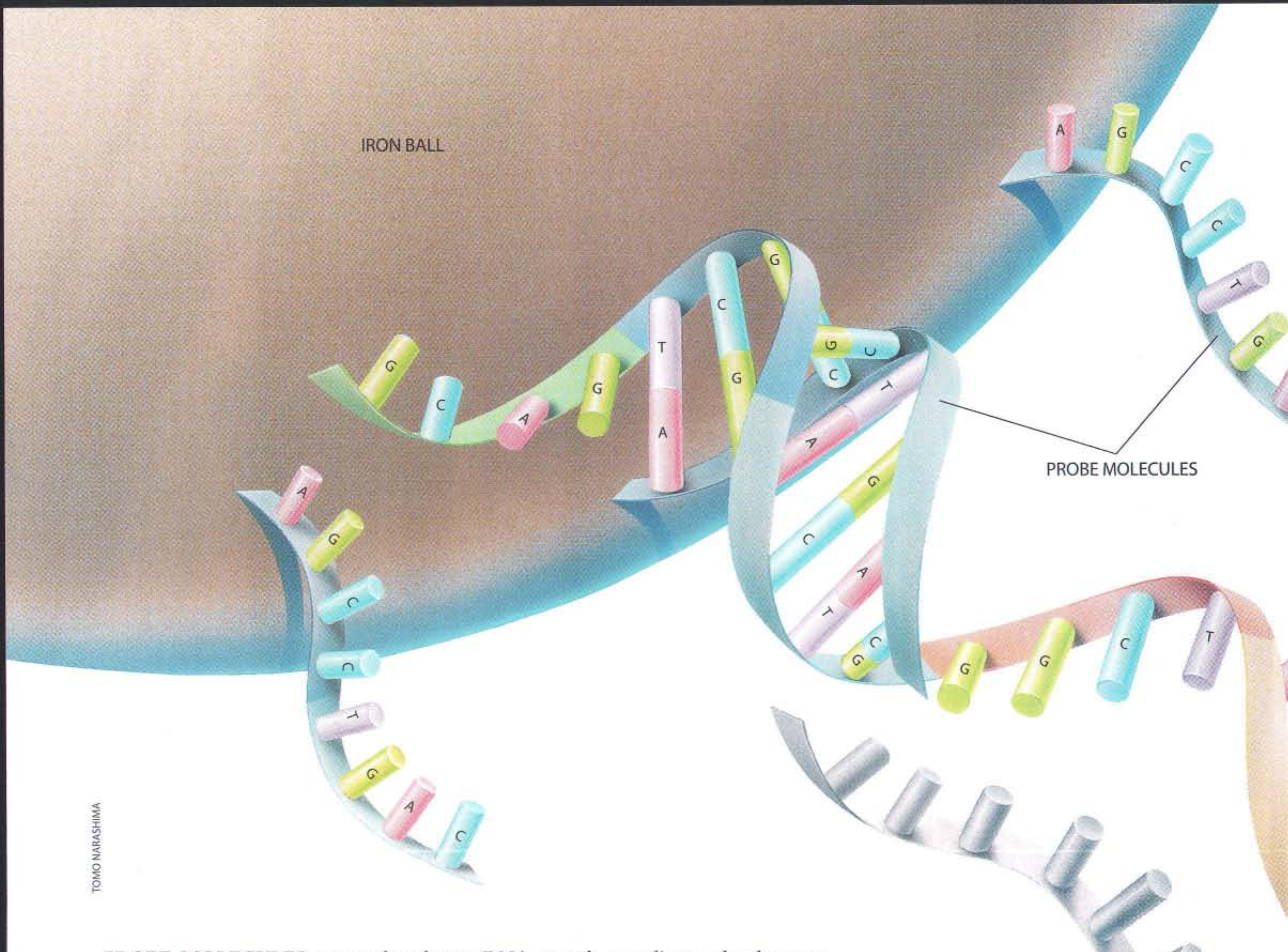


TOMO NARASHIMA



For the map on page 36, there is only one Hamiltonian path, and it goes through Atlanta, Boston, Chicago and Detroit, in that order. Thus, the molecule encoding the solution will have the sequence *GCAGTCGGACTGGGCTATGTCCGA*.

Unfortunately, although I held the solution in my hand, I also held about 100 trillion molecules that encoded paths that were not Hamiltonian. These had to be eliminated. To weed out molecules that did not both begin with the start city and terminate with the end city, I relied on the polymerase chain reaction (PCR). This important technique requires many copies of two short pieces of DNA as primers to signal the DNA polymerase to start its Watson-Crick replication. The primers used were the last name of the start city (*GCAG* for



PROBE MOLECULES are used to locate DNA strands encoding paths that pass through the intermediate cities (Boston and Chicago). Probe molecules containing the complementary DNA name of Boston (*AGCCTGAC*) are attached to an iron ball suspended in liquid. Because of Watson-Crick affinity, the probes capture DNA strands that contain Boston's name (*TCGGACTG*). Strands missing Boston's name are then discarded. The process is repeated with probe molecules encoding the complementary DNA name of Chicago. When all the computational steps are completed, the strands left will be those that encode the solution *GCAGTCGGACTGGGCTATGTCCGA*.

Atlanta) and the Watson-Crick complement of the first name of the end city (*GGCT* for Detroit). These two primers worked in concert: the first alerted DNA polymerase to copy complements of sequences that had the right start city, and the second initiated the duplication of molecules that encoded the correct end city.

PCR proceeds through thermocycling, repeatedly raising and lowering the temperature of the mixture in the test tube. Warm conditions encourage the DNA polymerase to begin duplicating; a hotter environment causes the resulting annealed strands to split from their double-helix structure, enabling subsequent replication of the individual pieces.

The result was that molecules with both the right start and end cities were reproduced at an exponential rate. In contrast, molecules that encoded the right start city but an incorrect end city, or vice versa, were duplicated in a much slower, linear fashion. DNA sequences that had neither the right start nor end were not duplicated at all. Thus, by taking a small amount of the mixture after the PCR was completed, I obtained a solution containing many copies of the molecules that had both the right start and end cities, but few if any molecules that did not meet this criterion. Thus, step 2a of the algorithm was complete.

Next, I used gel electrophoresis to identify those molecules that had the right length (in the example on page 36, a length of 24). All other molecules were discarded. This completed step 2b of the algorithm.

To check the remaining sequences for whether their paths passed through all the intermediary cities, I took advantage of Watson-Crick annealing in a procedure called affinity separation. This process uses multiple copies of a DNA "probe" molecule that encodes the complementary name of a particular city (for example, Boston). These probes are attached to microscopic iron balls, each approximately one micron in diameter.

I suspended the balls in the tube containing the remaining molecules under conditions that encouraged Watson-Crick pairing. Only those molecules that contained the desired city's name (Boston) would anneal to the probes. Then I placed a magnet against the wall of the

test tube to attract and hold the metal balls to the side while I poured out the liquid phase containing molecules that did not have the desired city's name.

I then added new solvent and removed the magnet in order to resuspend the balls. Raising the temperature of the mixture caused the molecules to break free from the probes and redissolve in the liquid. Next, I reapplied the magnet to attract the balls again to the side of the test tube, but this time without any molecules attached. The liquid, which now contained the desired DNA strands (in the example, encoding paths that went through Boston), could then be poured into a new tube for further screening. The process was repeated for the remaining intermediary cities (Chicago, in this case). This iterative procedure, which took me an entire day to complete in the lab, was the most tedious part of the experiment.

At the conclusion of the affinity separations, step 2c of the algorithm was over, and I knew that the DNA molecules left in the tube should be precisely those encoding Hamiltonian paths. Hence, if the tube contained any DNA at all, I could conclude that a Hamiltonian path existed in the graph. No DNA would indicate that no such path existed. Fortunately, to make this determination I could use an additional PCR step, followed by another gel-electrophoresis operation. To my delight, the final analysis revealed that the molecules that remained did indeed encode the desired Hamiltonian path. After seven days in the lab, the first DNA computation was complete.

A New Field Emerges

What about the future? It is clear that molecular computers have many attractive properties. They provide extremely dense information storage. For example, one gram of DNA, which when dry would occupy a volume of approximately one cubic centimeter, can store as much information as approximately one trillion CDs. They provide enormous parallelism. Even in the tiny experiment carried out in one fiftieth of a teaspoon of solution, approximately 10^{14} DNA flight numbers were simultaneously concatenated in about one second. It is not clear whether the fastest supercomputer available today could accomplish such a task so quickly.

Molecular computers also have the potential for extraordinary energy efficiency. In principle, one joule is sufficient for approximately 2×10^{19} ligation operations. This is remarkable considering that the second law of thermodynamics dictates a theoretical maximum of 34×10^{19} (irreversible) operations per joule (at room temperature). Existing supercomputers are far less efficient, executing at most 10^9 operations per joule.

Experimental and theoretical scientists around the world are working to exploit these properties. Will they succeed in creating molecular computers that can compete with electronic computers? That remains to be seen. Huge financial and intellectual investments over half a century have made electronic computers the marvels of our age—they will be hard to beat.

But it would be shortsighted to view this research only in such practical terms. My experiment can be viewed as a manifestation of an emerging new area of science made possible by our rapidly developing ability to control the molecular world. Evidence of this new "molecular science" can be found in many places. For example, Gerald F. Joyce of Scripps Research Institute in La Jolla, Calif., "breeds" trillions of RNA molecules, generation after generation, until "champion" molecules evolve that have the catalytic properties he seeks [see "Directed Molecular Evolution," by Gerald F. Joyce; SCIENTIFIC AMERICAN, December 1992]. Julius Rebek, Jr., of the Massachusetts Institute of Technology creates molecules that can reproduce—informing us about how life on the earth may have arisen [see "Synthetic Self-Replicating Molecules," by Julius Rebek, Jr.; SCIENTIFIC AMERICAN, July 1994]. Stimulated by research on DNA computation, Erik Winfree of the California Institute of Technology synthesizes "intelligent" molecular complexes that can be "programmed" to assemble themselves into predetermined structures of arbitrary complexity. There are many other examples. It is the enormous potential of this new area that we should focus on and nurture.

For me, it is enough just to know that computation with DNA is possible. In the past half-century, biology and computer science have blossomed, and there can be little doubt that they will be central to our scientific and economic progress in the new millennium. But biology and computer science—life and computation—are related. I am confident that at their interface great discoveries await those who seek them. SA

The Author

LEONARD M. ADLEMAN received a Ph.D. in computer science in 1976 from the University of California, Berkeley. In 1977 he joined the faculty in the mathematics department at the Massachusetts Institute of Technology, where he specialized in algorithmic number theory and was one of the inventors of the RSA public-key cryptosystem. (The "A" in RSA stands for "Adleman.") Soon after joining the computer science faculty at the University of Southern California, he was "implicated" in the emergence of computer viruses. He is a member of the National Academy of Engineering.

Further Reading

MOLECULAR COMPUTATION OF SOLUTIONS TO COMBINATORIAL PROBLEMS. Leonard M. Adleman in *Science,* Vol. 266, pages 1021–1024; November 11, 1994.

ON THE PATH TO COMPUTATION WITH DNA. David K. Gifford in *Science,* Vol. 266, pages 993–994; November 11, 1994.

DNA SOLUTION OF HARD COMPUTATIONAL PROBLEMS. Richard J. Lipton in *Science,* Vol. 268, pages 542–545; April 28, 1995.

Additional information on DNA computing can be found at http://users.aol.com/ibrandt/dna_computer.html on the World Wide Web.

Monitoring and Controlling Debris in Space

by Nicholas L. Johnson

Since the space age began four decades ago, rockets have lifted more than 20,000 metric tons of material into orbit. Today 4,500 tons remain in the form of nearly 10,000 "resident space objects," only 5 percent of which are functioning spacecraft. These objects are just the large ones that military radars and telescopes can track. Of increasing interest to spacecraft operators are the millions of smaller, untrackable scraps scattered into orbits throughout near-Earth space, from only a few hundred kilometers to more than 40,000 kilometers (25,000 miles) above the surface of the planet.

If Earth's tiny attendants moved like the hordes of miniature moons around Jupiter or Saturn, they would be a thing of beauty. The rings of the giant planets are finely orchestrated; their constituent rocks and chunks of ice orbit in well-behaved patterns, and collisions between them occur at gentle velocities. But Earth's artificial satellites resemble angry bees around a beehive, seeming to move randomly in all directions. The population density of satellites is fairly low; the region around Earth is still a vacuum by any terrestrial standard. But the haphazard motions of the swarm lead to huge relative velocities when objects accidentally collide. A collision with a one-centimeter pebble can destroy a spacecraft. Even a single one-millimeter grain could wreck a mission.

The extraterrestrial refuse comes in many forms: dead spacecraft, discarded rocket bodies, launch- and mission-related castoffs, remnants of satellite breakups, solid-rocket exhaust, frayed surface materials and even droplets from leaking nuclear reactors [*see* a *in top illustration on page 44*].

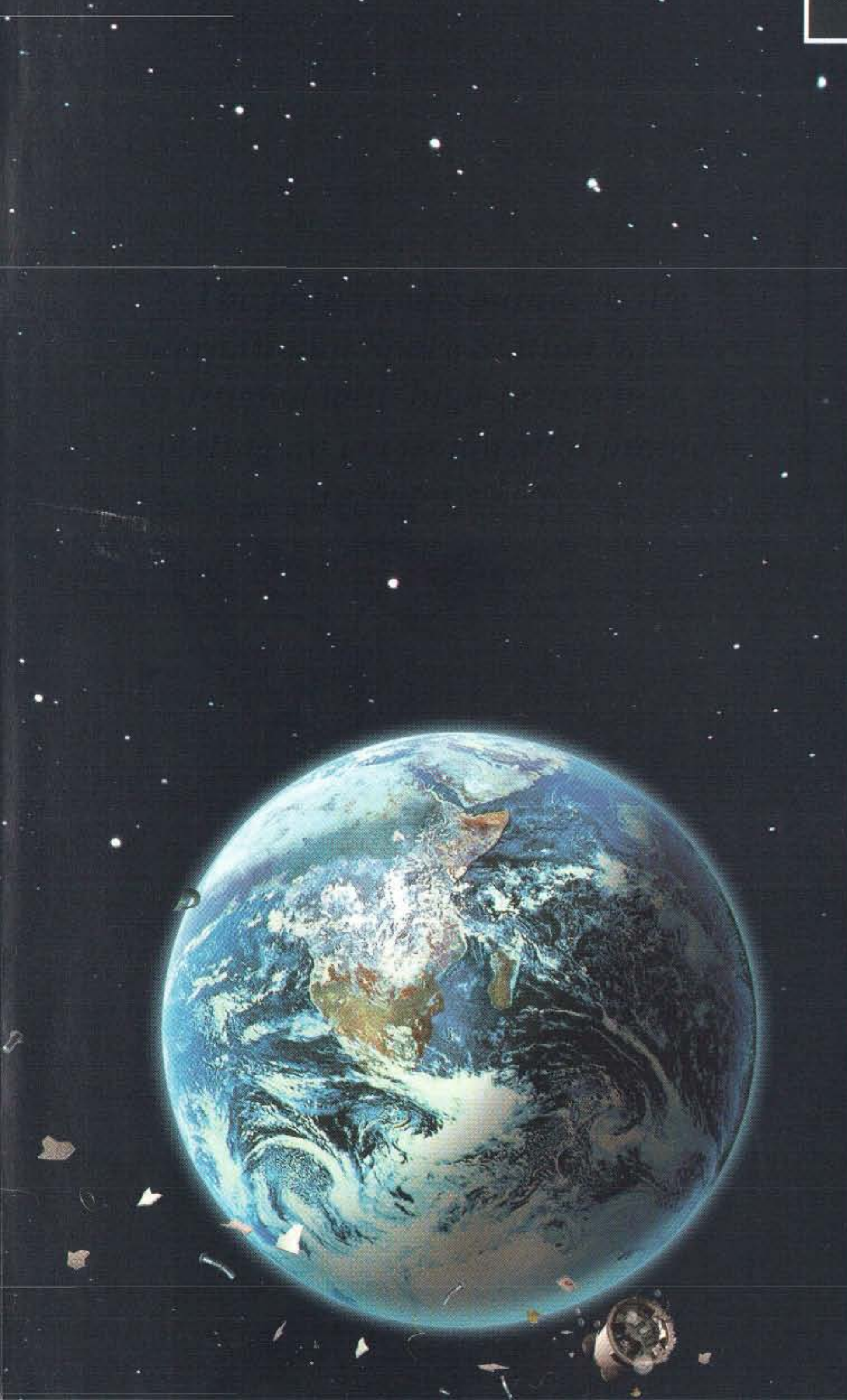
Although more than 4,800 spacecraft have been placed into orbit, only about 2,400 remain there; the rest have reentered Earth's atmosphere. Of the surviving spacecraft, 75 percent have completed their missions and been abandoned. Most range in mass from one kilogram to 20 tons, although the Russian Mir space station now exceeds 115 tons. The oldest and one of the smallest is America's second satellite, Vanguard 1, launched on March 17, 1958, but functional for only six years.

In addition to launching a spacecraft, most space missions also leave one or more empty rocket stages in orbit. For example, Japan's weather satellite, Himawari 3, launched in 1984, discarded three stages: one in a low Earth orbit that varied from 170 to 535 kilometers in altitude; one near the satellite's destination, the circular geosynchronous orbit at 35,785 kilometers; and one in an intermediate, highly elliptical orbit from 175 to 36,720 kilometers. Two of these rocket bodies have since fallen back to Earth, one in 1984 and the other in 1994. But by 1998 about 1,500 useless upper stages were still circling overhead.

For the first quarter century of the space age, spacecraft designers paid little attention to the environmental consequences of their actions. In addition to the abandoned spacecraft and leftover rockets, small components were routinely ejected into space. Explosive bolts, clamp bands and springs were released when satellites separated from the rockets that launched them. Many spacecraft also threw off sensor covers or attitude-control devices. Some Russian space missions spawned more than 60 distinct objects in various orbits. Perhaps the largest piece of discarded equipment,

SLIM FILMS

The path from Sputnik to the International Space Station has been littered with high-tech refuse, creating an environmental problem in outer space



ITT KAMAN SCIENCES CORPORATION

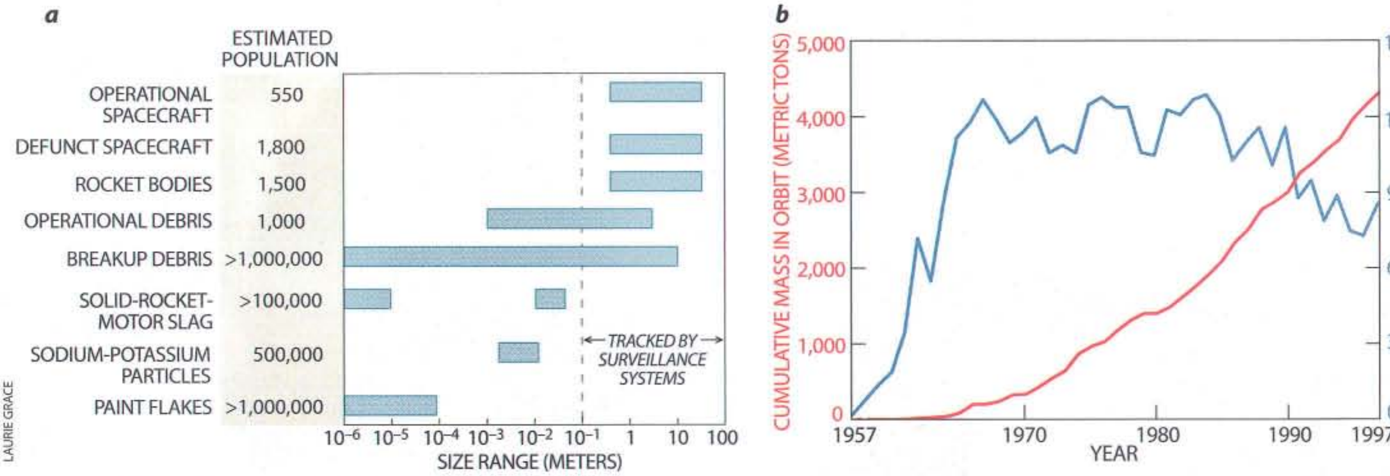
NUTS, BOLTS, CLAMPS AND WIRES are among the orbital flotsam from the breakup of an old rocket, as depicted in this artist's conception. Over time, the debris will scatter, reducing its density. Military radars and telescopes track the larger pieces (*inset, above*). White dots represent these objects (not to scale with Earth).

with a diameter of four meters and a mass of 300 kilograms, is the upper half of the European Space Agency's SPELDA (Structure Porteuse Externe pour Lancements Doubles Ariane), a device used for launching multiple satellites with a single Ariane rocket.

Human spaceflights, too, have thrown garbage overboard. In 1965, during the first space walk by an American, *Gemini 4* astronaut Edward White lost a glove; going into its own orbit at 28,000 kilometers per hour, it was potentially the most lethal article of clothing ever made by (and worn by) human hands. More than 200 objects, most in the form of garbage bags, drifted away from Mir during its first decade in orbit. Fortunately, because astronauts and cosmonauts orbit at fairly low altitudes, from 250 to 500 kilometers, the things they misplace soon fall into the atmosphere and burn up. The famous glove reentered the atmosphere within a month.

Breaking Up Isn't Hard to Do

The bigger problem comes from robotic missions at higher altitudes, where debris dawdles. Mission-related rubbish accounts for more than 1,000 known objects. Among them: 80 clumps of needles released in May 1963 as part of a U.S. Department of Defense telecommunications experiment. The radiation pressure exerted by sunlight was to have pushed the tiny needles—all 400 million of them—out of orbit, but a deployment malfunction caused them to



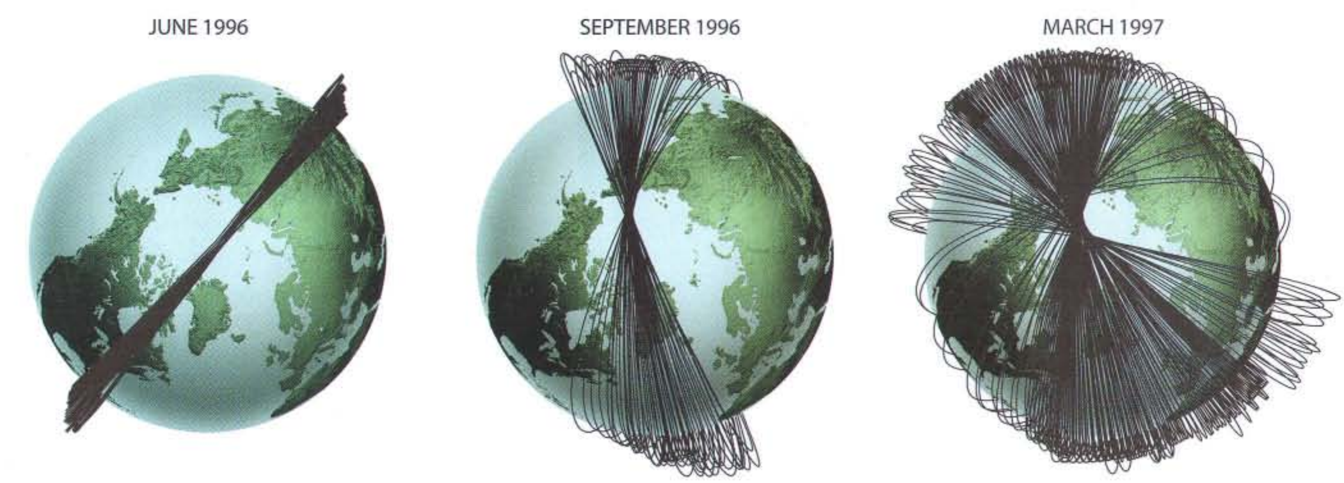
EARTH'S ARTIFICIAL SATELLITES, broadly defined as anything put into orbit either intentionally or not, come in various types and size ranges, from paint particles a thousandth of a millimeter across to the Mir space station, 30 meters long (*a*). Fewer new space missions are launched today than in the past, but the cumulative mass of satellites has continued to climb because modern spacecraft are larger (*b*). Military surveillance systems can spot only those satellites larger than 10 centimeters to one meter (depending on the orbit). Their counts have risen every year except when solar activity has been increasing (*c*). The density of these trackable objects has three closely spaced peaks at 850-, 1,000- and 1,500-kilometer altitudes, as well as smaller peaks at semisynchronous orbit (20,000 kilometers) and geosynchronous orbit (36,000 kilometers)—the most popular destinations (*d*).

bunch together and loiter in orbits up to 6,000 kilometers above Earth's surface.

By far the greatest source of space debris larger than 0.1 millimeter is the breakup of satellites and rockets. Since 1961 more than 150 satellites have blown up or fallen apart, either accidentally or deliberately—scattering more than 10,000 fragments large enough to be tracked. The most environmentally damaging fragmentations have been explosions of derelict rocket bodies with fuel left on board. Detonation, presumably caused by either overpressurization or ignition of the residual propellant, has occurred as soon as a few hours or as late as 23 years after launch, and nearly every booster type is vulnerable.

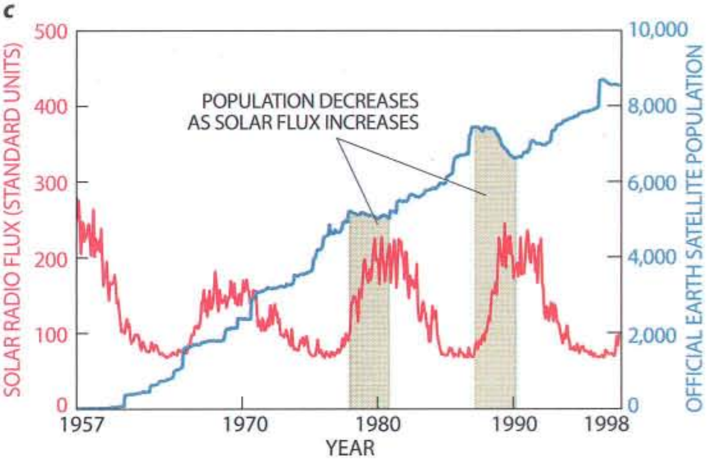
The upper stage of a Pegasus rocket launched in 1994 broke up on June 3, 1996, creating the largest debris cloud on record—more than 700 objects large enough to be tracked, in orbits from 250 to 2,500 kilometers in altitude [*see illustration below*]. This single event instantly doubled the official collision hazard to the Hubble Space Telescope, which orbits just 25 kilometers below. Further radar observations revealed an estimated 300,000 pieces of debris larger than four millimeters, big enough to damage most spacecraft. During the second Hubble servicing mission in February 1997, the space shuttle *Discovery* had to maneuver away from a piece of Pegasus debris that was projected to come within 1.5 kilometers. The astronauts also noticed pits on Hubble equipment and a hole in one of its antennas, which had apparently been caused by a collision with space debris before 1993.

Fifty satellites pose a special threat: they contain radioactive materials, either in nuclear reactors or in radioisotope thermoelectric generators. In 1978 a nuclear-powered satellite, the Soviet Kosmos 954, accidentally crash-landed in northern Canada with its 30 kilograms of enriched uranium. Soviet planners designed subsequent spacecraft to eject their fuel cores at the end of their

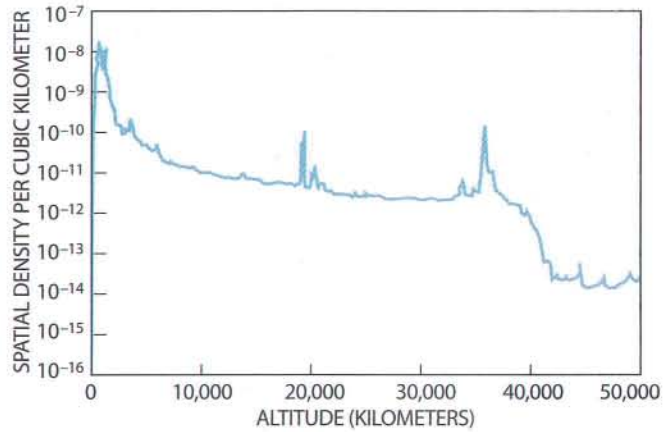


DEBRIS ORBITS disperse soon after a satellite breakup. Three weeks after a Pegasus rocket fell apart in June 1996, the debris was still concentrated in a tight band (*left*). After three months, the slight asymmetry of Earth's gravitational field had already started to amplify differences in the initial orbits (*center*). After nine months, the debris was spread out in essentially random order (*right*). In addition, though not depicted here, the objects initially orbited in a pack and slowly fell out of sync. Debris specialists must take this process into account when assessing the risk to spacecraft such as the shuttle. Initially the shuttle may avoid the concentrated debris. Later, the risk of impact from different directions must be evaluated statistically.

c



d



missions. Ejecting the fuel allows it to burn up if it reenters the atmosphere, rather than hit the ground as one lump.

This program ended altogether in 1988, and no nuclear reactors have been operated in orbit since then, but the problem has not gone away. During a 1989 experiment sponsored by the National Aeronautics and Space Administration, the Jet Propulsion Laboratory's Goldstone radar in southern California detected a large cloud of sodium-potassium droplets (leaking reactor coolant from an earlier ejection of a fuel core). Later observations by the Massachusetts Institute of Technology's Haystack radar confirmed a huge number of sodium-potassium spheres (perhaps 70,000), each about one centimeter across, near the 900-kilometer altitude where the reactors and cores were retired.

Objects smaller than 0.1 millimeter are not as hazardous as the larger junk, but their sheer quantity regularly inflicts minor damage on spacecraft. Numerically the most abundant debris is solid-rocket exhaust. Even the proper operation of a solid-rocket motor can yield a colossal number of micron-size aluminum oxide particles (up to 10^{20} of them), as well as centimeter-size slag. Although the use of solid-rocket motors in space has declined during the past 10 years, the total amount of debris they generate annually has actually risen because heavy, modern spacecraft require larger motors, which expel more exhaust.

Small particles in orbit also result from the long-term, natural degradation of spacecraft materials. Millions of tiny paint particles now litter near-Earth space. Space scientists have found that trails of minute flakes accompany many, if not most, aging spacecraft and rocket bodies. (Because of their relative orbital motions, these trails precede, rather than follow, the spacecraft: as the debris falls toward Earth, its speed increases.) Most of the detritus is invisible to remote sensors, but occasionally fragments from thermal blankets and carbon-based components are large enough to be detected. NASA's Cosmic Background Explorer satellite, for example, has inexplicably discarded at least 80 objects.

Exposing Satellites

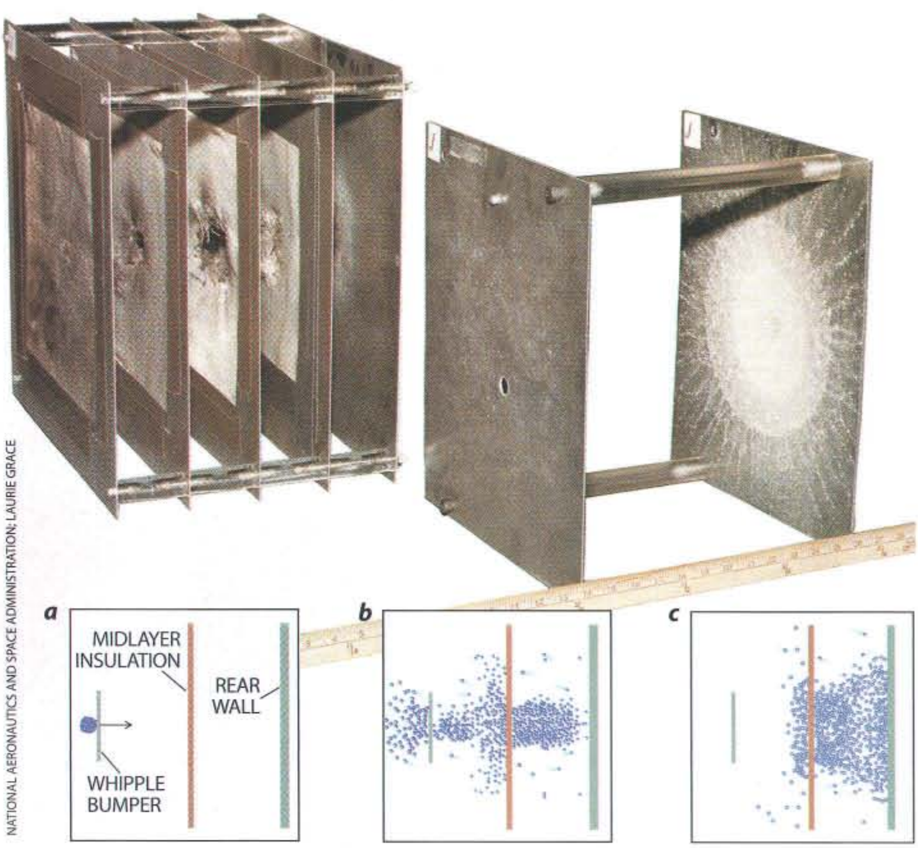
Although some futurists in the 1960s and 1970s speculated on the rising number of objects in orbit, not until the early 1980s did a scientific discipline emerge to address the issue. NASA, as early as 1966, evaluated the orbital collision hazards for human spaceflight, but these calculations considered only the objects that could be tracked. No means were then available to ascertain the number of smaller particles.

The larger pieces of space debris are monitored by the same tracking systems the superpowers built during the cold war to watch for missile attacks and enemy spy satellites. The Space Surveillance Network in the U.S. and the Space Surveillance System in the former Soviet Union maintain the official catalogue of some 10,000 resident space objects. Fifty radar, optical and electro-optical sensors collect an average of 150,000 observations a day to track these objects. They can spot scraps as small as 10 centimeters at lower altitudes and one meter at geosynchronous altitudes.

The spatial density of this debris depends on altitude, with peaks near 850, 1,000 and 1,500 kilometers. At these altitudes, there is an average of one object per 100 million cubic kilometers. Above 1,500 kilometers the density decreases with altitude, except for sharp peaks near semisynchronous (20,000 kilometers) and geosynchronous (36,000 kilometers) altitudes [*see* d *in illustration above*]. The clustering in particular orbits reflects the designs of the various families of satellites and rockets. Because the junk comes from satellites, it is densest in the regions where satellites tend to go, thereby amplifying the danger to functioning spacecraft.

Objects smaller than 10 centimeters escape the attention of the satellite trackers. They are too dim for the telescopes and too small for the radar wavelength used by the surveillance systems. Until 1984, these smaller objects went undetected. Since then, scientists have estimated their numbers by statistically sampling the sky over hundreds of hours every year. Researchers have used scientific radars, such as Goldstone, a German radio telescope and the huge Arecibo dish in Puerto Rico, in bistatic mode—transmitting signals from one dish and receiving the satellite reflections on a nearby antenna—to detect objects as small as two millimeters. The Haystack and Haystack Auxiliary radars, parts of a joint NASA–Department of Defense project, can detect objects nearly as small. To calibrate the sensors used to detect small debris, in 1994 and 1995 the space shuttle deployed special targets in the form of spheres and needles.

To study even smaller objects, researchers must inspect the battered surfaces of spacecraft that astronauts have pulled into the shuttle cargo bay and brought back to Earth. NASA's Long Duration Exposure Facility deliberately turned the other cheek, as it were, to space debris from 1984 to 1990; it was struck by tens of thousands of artificial shards, as well as natural meteoroids from comets and asteroids. Other sitting ducks have been the European Retrievable Carrier, Japan's Space Flyer Unit, components from Hubble and the Solar Maximum Mission and, of course, the space shuttle itself. Because of the shuttle's limited range, all these targets were restricted to altitudes below 620 kilometers. For higher altitudes, researchers



NATIONAL AERONAUTICS AND SPACE ADMINISTRATION; LAURIE GRACE

SHIELDS should protect key components of the International Space Station from most objects too small to track but large enough to puncture station walls. These shields, unlike the "force fields" of science fiction, are barriers mounted on the spacecraft. As the object approaches, it first encounters a sheet of aluminum (typically two millimeters thick), known as the Whipple bumper, which causes the projectile to shatter (*a*). The fragments are slowed by one or more layers of Kevlar (*b*). Finally, the fragments bounce off the spacecraft wall (*c*).

must rely on theoretical models. The short orbital lifetime of small particles at low altitudes implies that large reservoirs of them exist at higher altitudes.

In all, researchers estimate there are more than 100,000 objects between one and 10 centimeters in Earth orbit, along with tens of millions between one millimeter and one centimeter. In these size ranges and larger, artificial objects far outnumber natural meteoroids. The meteoroid and orbital debris populations are comparable in the 0.01- to 1.0-millimeter size regime; at very small sizes, orbital debris dominates again.

Births and Deaths

For the past two decades, the tracked satellite population has grown at an average rate of roughly 175 additional objects each year, for an overall increase of 70 percent. More than one quarter of this growth has been due to satellite breakups and the remainder to new missions. Since 1957 spacefaring nations have launched an average of 120 new spacecraft each year. Over the past decade, however, the activity has calmed down. It peaked at 129 missions in 1984 and plummeted to 73 in 1996, the lowest figure since 1963 [*see* b *in top illustration on page 44*]. The launch rate rebounded to 86 last year, largely because of the 10 launches for new communications satellite constellations (Iridium and Orbcom).

Much of the general decrease has occurred in the former Soviet Union. Russia had 28 space missions in 1997: 21 for domestic purposes, seven for foreign commercial firms. A decade earlier the U.S.S.R. launched 95 missions, all to support domestic programs. Because the shrinkage of Russian space activity has mostly affected low-altitude flights, however, it has had little effect on the long-term satellite population.

On the other side of the equation, three courses of action remove satellites. Satellite operators can deliberately steer their spacecraft into the atmosphere; the space shuttle can pluck satellites out of orbit; and satellite orbits can spiral downward naturally. The first two actions, though helpful, have little effect overall, because they typically apply only to objects in low orbits, which would have been removed naturally anyway.

The third—natural orbital decay—exerts the primary cleansing force. Although we usually think of outer space as airless, Earth's atmosphere actually peters out fairly slowly. There is still enough air resistance in low Earth orbit to generate friction against fast-moving spacecraft. This drag force is self-reinforcing: as satellites lose energy to drag, they fall inward and speed up, which increases the drag, causing them to lose energy faster. Eventually, satellites enter the denser parts of the atmosphere and burn up partially or wholly in a fiery streak.

This natural decay is more pronounced below 600 kilometers' altitude but is discernible as high as 1,500 kilometers. At the lower altitudes, satellites last only a few years unless they fire their rockets to compensate for the drag. During periods of higher-energy emissions from the sun, which recur on an 11-year cycle, Earth's atmosphere receives more heat and expands, increasing the atmospheric drag. Around the time of the last solar maximum in 1989–1990, a record number of catalogued satellites fell back to Earth—three a day, three times the average rate—sweeping more than 560 tons from orbit in a year's time [*see* c *in illustration on page 45*]. The Salyut 7 space station, the predecessor of Mir, fell victim to this activity, reentering in early 1991. The previous solar maximum in 1979–1980 brought down the U.S. Skylab space station ahead of schedule. Welcome though the natural garbage collection has been, it has not prevented the total satellite population from increasing. And atmospheric drag is negligible for satellites in higher orbits. Therefore, we cannot rely entirely on natural processes to solve the debris problem.

The consequences of the growth in Earth's satellite population are varied. The space shuttle occasionally executes an evasive maneuver to dodge large, derelict satellites, and an average of one in eight of its windows must be replaced after each mission because of pits gouged by hypervelocity dust. Increasingly, mission planners must consider the risk of impacts when they choose the orientation in which the shuttle flies. Debris is most likely to strike in the orbital plane 30 to 45 degrees left and right of the direction of travel. If there is no compelling need to fly in a different orientation, the astronauts point the most sen-

sitive shuttle surfaces away from these threatening directions.

In July 1996 the first recognized accidental collision of two catalogued satellites occurred. The functioning French military spacecraft Cerise was disabled when a fragment from a European rocket body, which had exploded 10 years earlier, struck the spacecraft's attitude-control boom at a speed of nearly 15 kilometers per second (nearly 33,400 miles per hour). Though damaged, the spacecraft was able to continue its mission after heroic efforts by its controllers.

Unfortunately, all spacecraft remain vulnerable to objects from one to 10 centimeters in size—too small for the surveillance systems to see but large enough to puncture spacecraft walls. The Department of Defense is slowly moving to improve its space surveillance network to cover this range of sizes, but these systems are intended for human spaceflight; they offer little protection for most satellites.

Vacuum Cleaners

Most spacecraft are still at little risk of impacts during their operational lifetimes, but the environment in the future is likely to be less benign. The Cerise collision generated little extra debris, but such a collision could create thousands of fragments large enough to shatter other satellites, which would create their own debris, and so on. This cascading phenomenon might become the dominant factor in the far distant evolution of Earth's satellite population.

To prevent such a scenario, spacefaring nations have begun working together over the past decade. The Inter-Agency Space Debris Coordination Committee currently includes representatives from the U.S., Russia, China, Japan, India, the European Space Agency, France, Britain and Germany. In 1994 the United Nations Scientific and Technical Subcommittee on the Peaceful Uses of Outer Space finally placed orbital debris on its agenda, with the goal of completing an assessment by 1999. Although everyone now recognizes the problem, the committees have yet to decide what should be done, who should police the situation, and how a balance should be struck between the risks of damage and the costs of mitigation and cleaning up.

Meanwhile individual agencies are striving to generate less garbage. NASA and Japan's principal space agency, NASDA, have used bolt catchers and special tethers to curtail the release of operational debris; dumped fuel and turned off electrical systems on dead spacecraft to prevent breakups; and recommended that new low-altitude satellites and rocket bodies be "de-orbited"—that is, steered into the atmosphere for incineration—no more than 25 years after completing their missions.

After the devastating breakup of the Pegasus in 1996, the designer and operator of the rocket, Orbital Sciences Corporation, redesigned the upper stage and implemented new preventive measures before missions resumed last December. Some of the new satellite-telephone networks plan to de-orbit satellites on retirement. At higher altitudes, the collisional dangers are fewer, but there is another problem: the lack of room in geosynchronous orbit. To keep dead satellites from occupying slots needed for working ones, satellite operators are now asked to give their spacecraft a proper burial in less crowded "graveyard" orbits. In January the U.S. government presented draft debris-mitigation standards to the aerospace industry for comment.

These procedures, however, do nothing about the junk already in space. Spacecraft designers are now encouraged to protect against collisions, particularly by the many objects up to one centimeter in size. Shields—that is, exterior barriers—can protect most spacecraft components. Not only do they increase reliability, but they reduce the amount of secondary debris in the event a collision does occur. The International Space Station will contain advanced-technology shields around habitable compartments, fuel lines, control gyroscopes and other sensitive areas [*see illustration on opposite page*]. Other components, such as solar panels, are impractical to shield; spacecraft planners must assume that they will slowly deteriorate over time because of small collisions.

Cleaning up the debris remains a technological and economic challenge. Using the shuttle to salvage debris is dangerous and impractical. Several researchers have devised imaginative schemes. NASA, the Department of Defense and the Department of Energy have studied one proposal, Project Orion, that would use ground-based lasers to remove small debris. These lasers are nothing like those in science fiction; even if they could blast a satellite into pieces, that would simply create more debris. Instead the lasers vaporize some satellite material, nudging it off course and eventually into the atmosphere. Other mechanisms are still on the drawing boards, including giant foam balls. A particle penetrating the foam would lose energy and fall back to Earth sooner.

Whereas future generations may be able to undo the shortsighted practices of the past, for now we must strive to prevent the uncontrolled growth of the satellite population—or else prepare ourselves for a lack of space in space. SA

The Author

NICHOLAS L. JOHNSON serves as the National Aeronautics and Space Administration's chief scientist for orbital debris at the Johnson Space Center in Houston. He is head of the space agency's delegation to the Inter-Agency Space Debris Coordination Committee, a co-chair of the International Academy of Astronautics's Subcommittee on Space Debris and a member of the American Institute of Aeronautics and Astronautics's Orbital Debris Committee on Standards. Johnson was a member of the National Research Council's space debris committee (1993–1994) and the International Space Station meteoroid/orbital debris risk management committee (1995–1996).

Further Reading

ARTIFICIAL SPACE DEBRIS. N. L. Johnson and D. S. McKnight. Orbit Books, 1991.

PRESERVATION OF NEAR EARTH SPACE FOR FUTURE GENERATIONS. Edited by J. A. Simpson. Cambridge University Press, 1994.

INTERAGENCY REPORT ON ORBITAL DEBRIS. Office of Science and Technology Policy, National Science Technology Council, November 1995. Available at http://www-sn.jsc.nasa.gov/debris/report95.html on the World Wide Web.

ORBITAL DEBRIS: A TECHNOLOGICAL ASSESSMENT. Committee on Space Debris, Aeronautics and Space Engineering Board, National Research Council. National Academy Press, 1995.

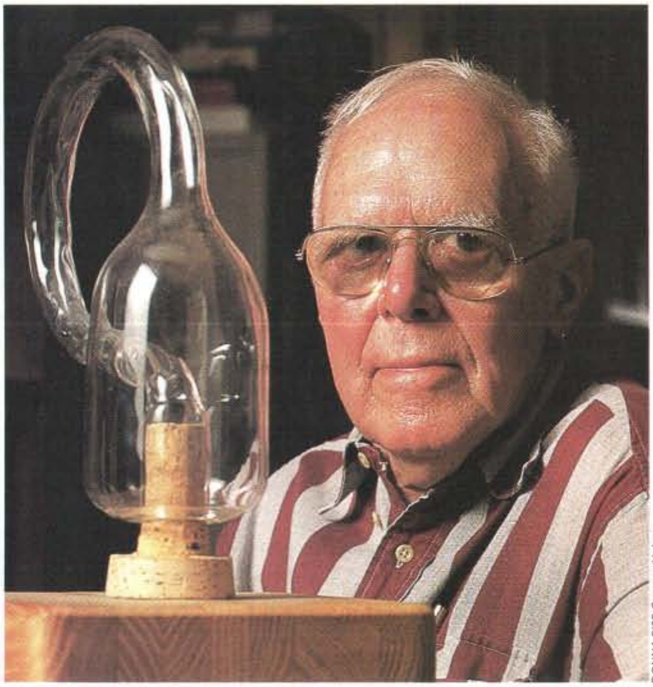
PROTECTING THE SPACE STATION FROM METEOROIDS AND ORBITAL DEBRIS. Committee on International Space Station Meteoroid/Orbital Debris Risk Management, Aeronautics and Space Engineering Board, National Research Council. National Academy Press, 1997.

A Quarter-Century of Recreational Mathematics

The author of Scientific American's *column "Mathematical Games" from 1956 to 1981 recounts 25 years of amusing puzzles and serious discoveries*

by Martin Gardner

"Amusement is one of the fields of applied math."
—William F. White,
A Scrapbook of Elementary Mathematics



MARTIN GARDNER continues to tackle mathematical puzzles at his home in Hendersonville, N.C. The 83-year-old writer poses next to a Klein bottle, an object that has just one surface: the bottle's inside and outside connect seamlessly.

My "Mathematical Games" column began in the December 1956 issue of *Scientific American* with an article on hexaflexagons. These curious structures, created by folding an ordinary strip of paper into a hexagon and then gluing the ends together, could be turned inside out repeatedly, revealing one or more hidden faces. The structures were invented in 1939 by a group of Princeton University graduate students. Hexaflexagons are fun to play with, but more important, they show the link between recreational puzzles and "serious" mathematics: one of their inventors was Richard Feynman, who went on to become one of the most famous theoretical physicists of the century.

At the time I started my column, only a few books on recreational mathematics were in print. The classic of the genre—*Mathematical Recreations and Essays*, written by the eminent English mathematician W. W. Rouse Ball in 1892—was available in a version updated by another legendary figure, the Canadian geometer H.S.M. Coxeter. Dover Publications had put out a translation from the French of *La Mathématique des Jeux* (*Mathematical Recreations*), by Belgian number theorist Maurice Kraitchik. But aside from a few other puzzle collections, that was about it.

Since then, there has been a remarkable explosion of books on the subject, many written by distinguished mathematicians. The authors include Ian Stewart, who now writes *Scientific American*'s "Mathematical Recreations" column; John H. Conway of Princeton University; Richard K. Guy of the University of Calgary; and Elwyn R. Berlekamp of the University of California at Berkeley. Articles on recreational mathematics also appear with increasing frequency in mathematical periodicals. The quarterly *Journal of Recreational Mathematics* began publication in 1968.

The line between entertaining math and serious math is a blurry one. Many professional mathematicians regard their work as a form of play, in the same way professional golfers or basketball stars might. In general, math is considered recreational if it has a playful aspect that can be understood and appreciated by nonmathematicians. Recreational math includes elementary problems with elegant, and at times surprising, solutions. It also encompasses mind-bending paradoxes, ingenious games, bewildering magic tricks and topological curiosities such as Möbius bands and Klein bottles. In fact, almost every branch of mathematics simpler than calculus has areas that can be considered recreational. (Some amusing examples are shown on the opposite page.)

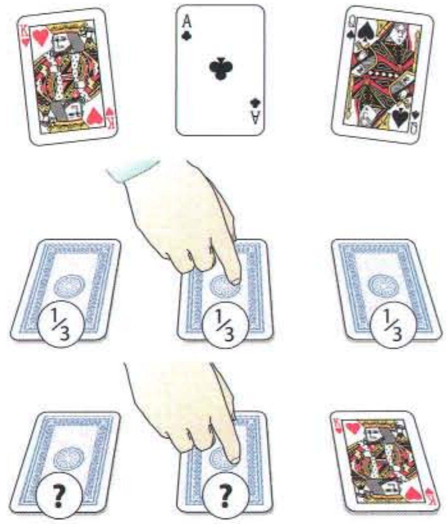
Ticktacktoe in the Classroom

The monthly magazine published by the National Council of Teachers of Mathematics, *Mathematics Teacher*, often carries articles on recreational topics. Most teachers, however, continue to

Four Puzzles from Martin Gardner

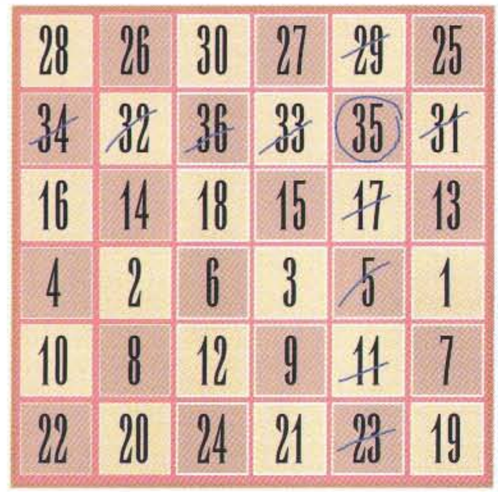
(The answers are on page 55.)

1



Mr. Jones, a cardsharp, puts three cards face down on a table. One of the cards is an ace; the other two are face cards. You place a finger on one of the cards, betting that this card is the ace. The probability that you've picked the ace is clearly $1/3$. Jones now secretly peeks at each card. Because there is only one ace among the three cards, at least one of the cards you *didn't* choose must be a face card. Jones turns over this card and shows it to you. What is the probability that your finger is now on the ace?

2



ILLUSTRATIONS BY IAN WORPOLE

The matrix of numbers above is a curious type of magic square. Circle any number in the matrix, then cross out all the numbers in the same column and row. Next, circle any number that has not been crossed out and again cross out the row and column containing that number. Continue in this way until you have circled six numbers.

Clearly, each number has been randomly selected. But no matter which numbers you pick, they always add up to the same sum. What is this sum? And, more important, why does this trick always work?

3

In the beginning God created the
1 2 3 4 5 6

heaven and the earth.
7 8 9 10

And the earth was without form,
11 12 13 14 15 16

and void; and darkness was upon the
17 18 19 20 21 22 23

face of the deep. And the Spirit of God
24 25 26 27 28 29 30 31 32

moved upon the face of the waters.
33 34 35 36 37 38 39

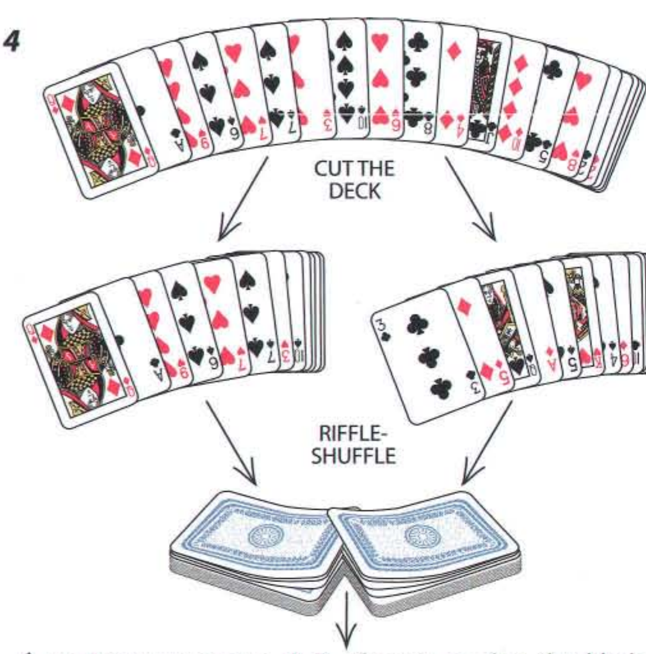
And God said, Let there be light:
40 41 42 43 44 45 46

and there was light.
47 48 49 50

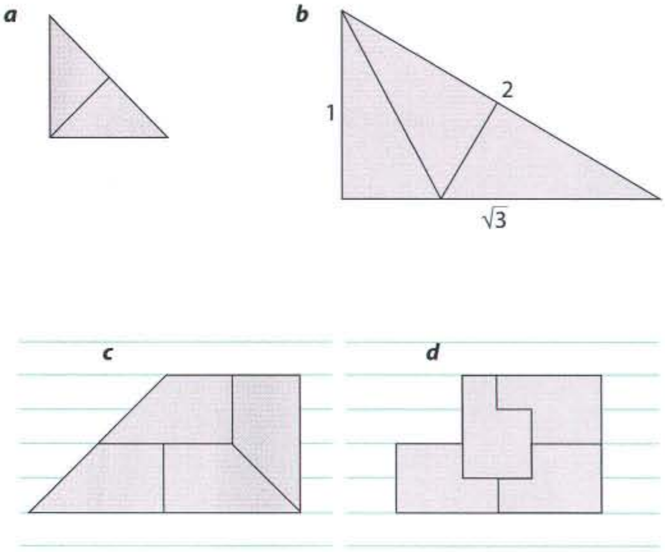
Printed above are the first three verses of Genesis in the King James Bible. Select any of the 10 words in the first verse: "In the beginning God created the heaven and the earth." Count the number of letters in the chosen word and call this number *x*. Then go to the word that is *x* words ahead. (For example, if you picked "in," go to "beginning.") Now count the number of letters in this word—call it *n*—then jump ahead another *n* words. Continue in this manner until your chain of words enters the third verse of Genesis.

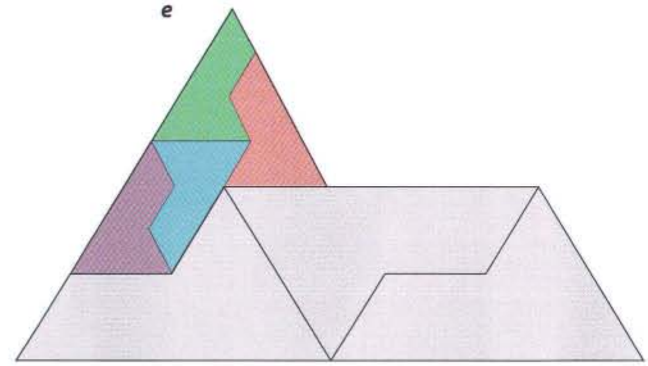
On what word does your count end? Is the answer happenstance or part of a divine plan?

4



A magician arranges a deck of cards so that the black and red cards alternate. She cuts the deck about in half, making sure that the bottom cards of each half are not the same color. Then she allows you to riffle-shuffle the two halves together, as thoroughly or carelessly as you please. When you're done, she picks the first two cards from the top of the deck. They are a black card and a red card (not necessarily in that order). The next two are also a black card and a red card. In fact, every succeeding pair of cards will include one of each color. How does she do it? Why doesn't shuffling the deck produce a random sequence?





IAN WORPOLE

LOW-ORDER REP-TILES fit together to make larger replicas of themselves. The isosceles right triangle (*a*) is a rep-2 figure: two such triangles form a larger triangle with the same shape. A rep-3 triangle (*b*) has angles of 30, 60 and 90 degrees. Other rep-tiles include a rep-4 quadrilateral (*c*) and a rep-4 hexagon (*d*). The sphinx (*e*) is the only known rep-4 pentagon.

ignore such material. For 40 years I have done my best to convince educators that recreational math should be incorporated into the standard curriculum. It should be regularly introduced as a way to interest young students in the wonders of mathematics. So far, though, movement in this direction has been glacial.

I have often told a story from my own high school years that illustrates the dilemma. One day during math study period, after I'd finished my regular assignment, I took out a fresh sheet of paper and tried to solve a problem that had intrigued me: whether the first player in a game of ticktacktoe can always win, given the right strategy. When my teacher saw me scribbling, she snatched the sheet away from me and said, "Mr. Gardner, when you're in my class I expect you to work on mathematics and nothing else."

The ticktacktoe problem would make a wonderful classroom exercise. It is a superb way to introduce students to combinatorial mathematics, game theory, symmetry and probability. Moreover, the game is part of every student's experience: Who has not, as a child, played ticktacktoe? Yet I know few mathematics teachers who have included such games in their lessons.

According to the 1997 yearbook of the mathematics teachers' council, the latest trend in math education is called "the new new math" to distinguish it from "the new math," which flopped so disastrously several decades ago. The newest teaching system involves dividing classes into small groups of students and instructing the groups to solve problems through cooperative reasoning. "Interactive learning," as it is called, is substituted for lecturing. Although there are some positive aspects of the new new math, I was struck by the fact that the yearbook had nothing to say about the value of recreational mathematics, which lends itself so well to cooperative problem solving.

Let me propose to teachers the following experiment. Ask each group of students to think of any three-digit number—let's call it ABC. Then ask the students to enter the sequence of digits twice into their calculators, forming the number ABCABC. For example, if the students thought of the number 237, they'd punch in the number 237,237. Tell the students that you have the psychic power to predict that if they divide ABCABC by 13 there will be no remainder. This will prove to be true. Now ask them to divide the result by 11. Again, there will be no remainder. Finally, ask them to divide by 7. Lo and behold, the original number ABC is now in the calculator's readout. The secret to the trick is simple: ABCABC = ABC × 1,001 = ABC × 7 × 11 × 13. (Like every other integer, 1,001 can be factored into a unique set of prime numbers.) I know of no better introduction to number theory and the properties of primes than asking students to explain why this trick always works.

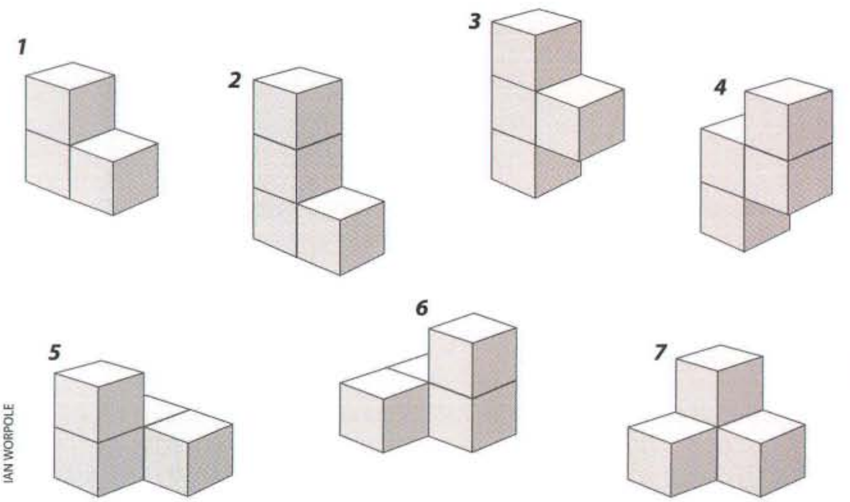
Polyominoes and Penrose Tiles

One of the great joys of writing the *Scientific American* column over 25 years was getting to know so many authentic mathematicians. I myself am little more than a journalist who loves mathematics and can write about it glibly. I took no math courses in college. My columns grew increasingly sophisticated as I learned more, but the key to the column's popularity was the fascinating material I was able to coax from some of the world's best mathematicians.

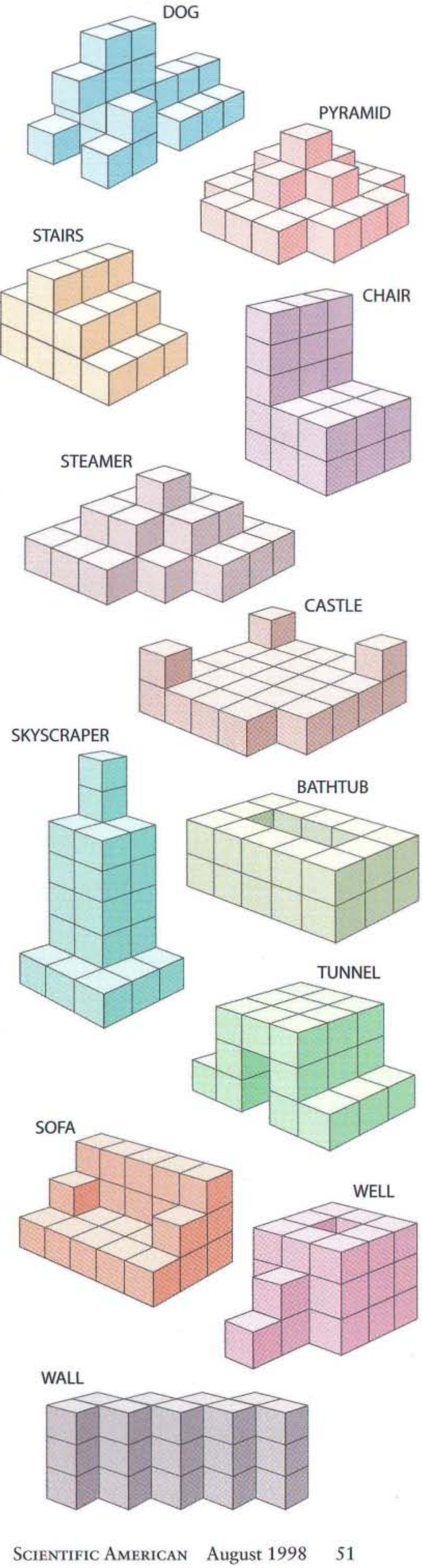
Solomon W. Golomb of the University of Southern California was one of the first to supply grist for the column. In the May 1957 issue I introduced his studies of polyominoes, shapes formed by joining identical squares along their edges. The domino—created from two such squares—can take only one shape, but the tromino, tetromino and pentomino can assume a variety of forms: Ls, Ts, squares and so forth. One of Golomb's early problems was to determine whether a specified set of polyominoes, snugly fitted together, could cover a checkerboard without missing any squares. The study of polyominoes soon evolved into a flourishing branch of recreational mathematics. Arthur C. Clarke, the science-fiction author, confessed that he had become a "pentomino addict" after he started playing with the deceptively simple figures.

Golomb also drew my attention to a class of figures he called "rep-tiles"—identical polygons that fit together to form larger replicas of themselves. One of them is the sphinx, an irregular pentagon whose shape is somewhat similar to that of the ancient Egyptian monument. When four identical sphinxes are joined in the right manner, they form a larger sphinx with the same shape as its components. The pattern of rep-tiles can expand infinitely: they tile the plane by making larger and larger replicas.

The late Piet Hein, Denmark's illustrious inventor and poet, became a



IAN WORPOLE



SOMA PIECES are irregular shapes formed by joining unit cubes at their faces (*above*). The seven pieces can be arranged in 240 ways to build the 3-by-3-by-3 Soma cube. The pieces can also be assembled to form all but one of the structures pictured at the right. Can you determine which structure is impossible to build? The answer is on page 55.

good friend through his contributions to "Mathematical Games." In the July 1957 issue, I wrote about a topological game he invented called Hex, which is played on a diamond-shaped board made of hexagons. Players place their markers on the hexagons and try to be the first to complete an unbroken chain from one side of the board to the other. The game has often been called John because it can be played on the hexagonal tiles of a bathroom floor.

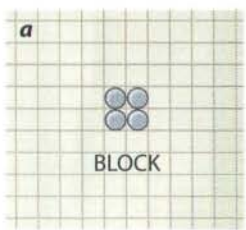
Hein also invented the Soma cube, which was the subject of several columns (September 1958, July 1969 and September 1972). The Soma cube consists of seven different polycubes, the three-dimensional analogues of polyominoes. They are created by joining identical cubes at their faces. The polycubes can be fitted together to form the Soma cube—in 240 ways, no less—as well as a whole panoply of Soma shapes: the pyramid, the bathtub, the dog and so on.

In 1970 the mathematician John Conway—one of the world's undisputed geniuses—came to see me and asked if I had a board for the ancient Oriental game of go. I did. Conway then demonstrated his now famous simulation game called Life. He placed several counters on the board's grid, then removed or added new counters according to three simple rules: each counter with two or three neighboring counters is allowed to remain; each counter with one or no neighbors, or four or more neighbors, is removed; and a new counter is added to each empty space adjacent to exactly three counters. By applying these rules repeatedly, an astonishing variety of forms can be created, including some that move across the board like insects. I described Life in the October 1970 column, and it became an instant hit among computer buffs. For many weeks afterward, business firms and research laboratories were almost shut down while Life enthusiasts experimented with Life forms on their computer screens.

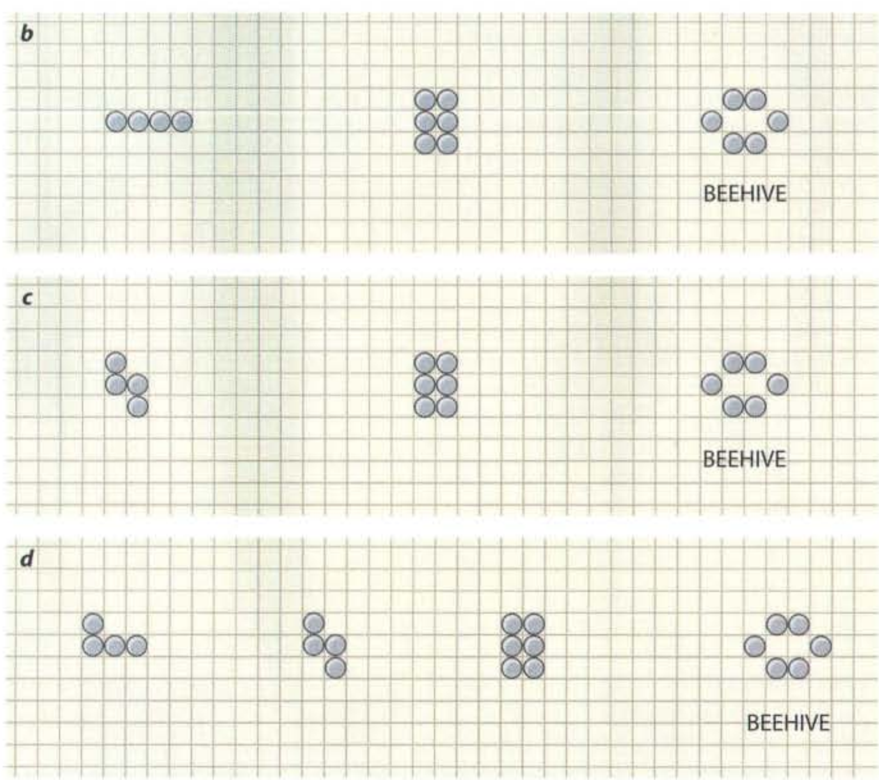
Conway later collaborated with fellow mathematicians Richard Guy and Elwyn Berlekamp on what I consider the greatest contribution to recreational mathematics in this century, a two-volume work called *Winning Ways* (1982). One of its hundreds of gems is a two-person game called Phutball, which can also be played on a go board. The Phutball is positioned at the center of the board, and the players take turns placing counters on the intersections of the grid lines. Players can move the Phutball by jumping it over the counters, which are removed from the board after they have been leapfrogged. The object of the game is to get the Phutball past the opposing side's goal line by building a chain of counters across the board. What makes the game distinctive is that, unlike checkers, chess, go or Hex, Phutball does not assign different game pieces to each side: the players use the same counters to build their chains. Consequently, any move made by one Phutball player can also be made by his or her opponent.

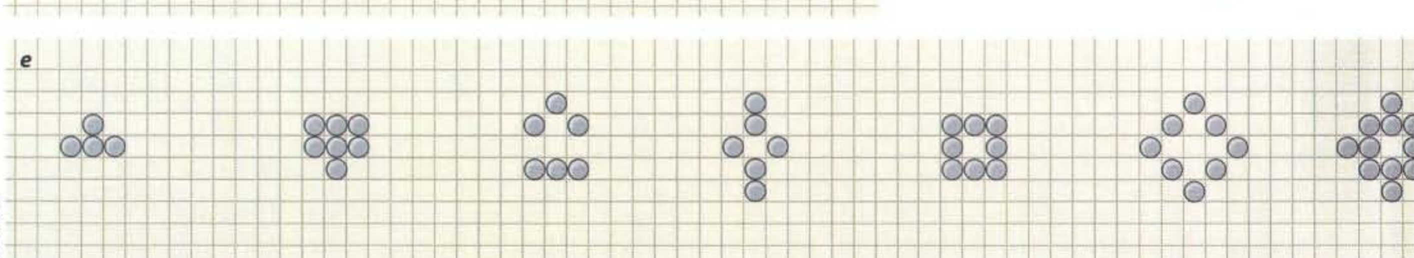
Other mathematicians who contributed ideas for the column include Frank Harary, now at New Mexico State University, who generalized the game of ticktacktoe. In Harary's version of the game, presented in the April 1979 issue, the goal was not to form a straight line of Xs or Os; instead players tried to be the first to arrange their Xs or Os in a specified polyomino, such as an L or a square. Ronald L. Rivest of the Massachusetts Institute of Technology allowed me to be the first to reveal—in the August 1977 column—the "public-key" cipher system that he co-invented. It was the first of a series of ciphers that revolutionized the field of cryptology. I also had the pleasure of presenting the mathematical art of Maurits C. Escher, which appeared on the cover of the April 1961 issue of *Scientific American,* as well as the nonperiodic tiling discovered by Roger Penrose, the British mathematical physicist famous for his work on relativity and black holes.

Penrose tiles are a marvelous example of how a discovery made solely for the fun of it can turn out to have an unexpected practical use. Penrose devised two kinds of shapes, "kites" and "darts," that cover the plane only in a nonperiodic way: no fundamental part of the pattern repeats itself. I explained the significance of the discovery in the January 1977 issue, which featured a pattern of Penrose tiles on its cover. A few years later a 3-D form of Penrose tiling became the basis for constructing a previously unknown type of molecular structure called a quasicrystal. Since then, physicists have written hundreds of research papers on quasicrystals and their unique thermal and vibrational properties. Although Penrose's idea started as a strictly recreational pursuit, it paved the way for an entirely new branch of solid-state physics.



IN THE GAME OF LIFE, forms evolve by following rules set by mathematician John H. Conway. If four "organisms" are initially arranged in a square block of cells (*a*), the Life form does not change. Three other initial patterns (*b*, *c* and *d*) evolve into the stable "beehive" form. The fifth pattern (*e*) evolves into the oscillating "traffic lights" figure, which alternates between vertical and horizontal rows.



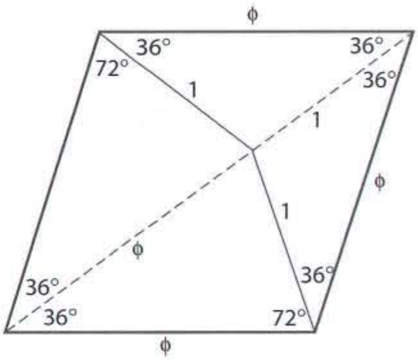


IAN WORPOLE

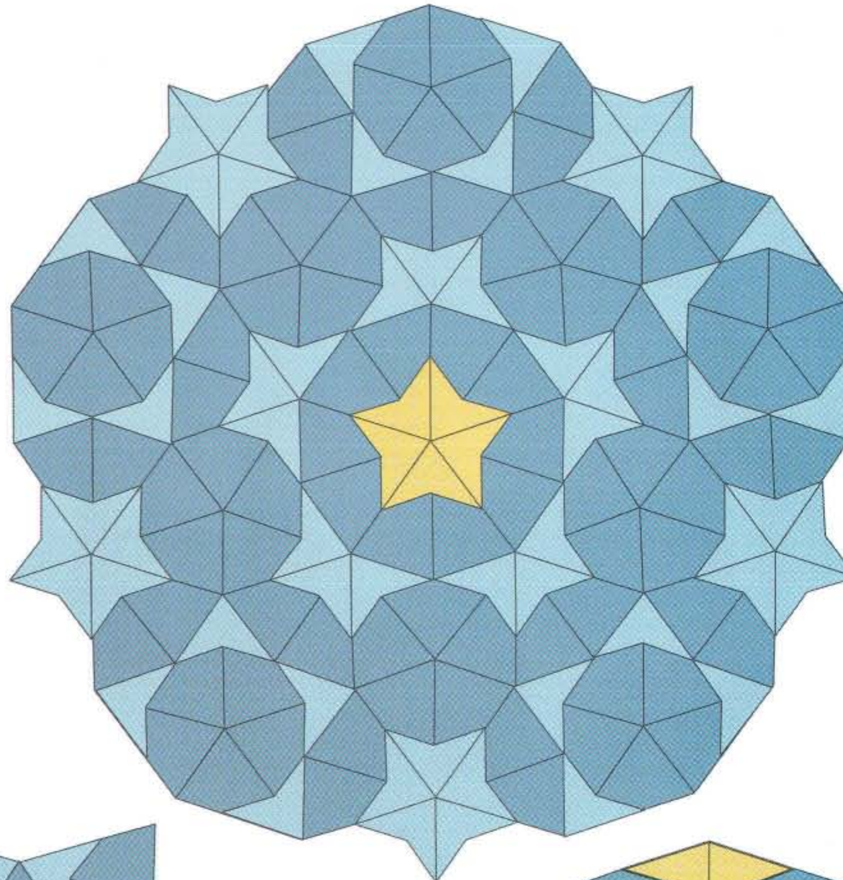
Leonardo's Flush Toilet

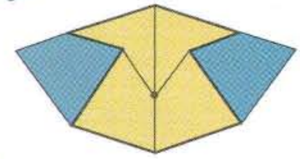
The two columns that generated the greatest number of letters were my April Fools' Day column and the one on Newcomb's paradox. The hoax column, which appeared in the April 1975 issue, purported to cover great breakthroughs in science and math. The startling discoveries included a refutation of relativity theory and the disclosure that Leonardo da Vinci had invented the flush toilet. The column also announced that the opening chess move of pawn to king's rook 4 was a certain game winner and that *e* raised to the power of $\pi \times \sqrt{163}$ was exactly equal to the integer 262,537,412,640,768,744. To my amazement, thousands of readers failed to recognize the column as a joke. Accompanying the text was a complicated map that I said required five colors to ensure that no two neighboring regions were colored the same. Hundreds of readers sent me copies of the map colored with only four colors, thus upholding the four-color theorem. Many readers said the task had taken days.

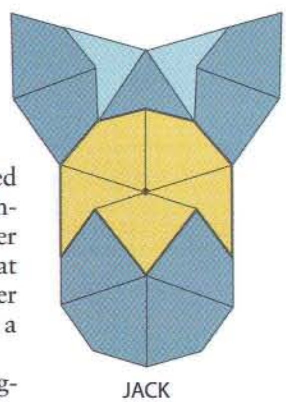
Newcomb's paradox is named after physicist William A. Newcomb, who originated the idea, but it was first described in a technical paper by Harvard University philosopher Robert Nozick. The paradox involves two closed boxes, A and B. Box A contains \$1,000. Box B contains either nothing or \$1 million. You have two choices: take only Box B or take both boxes. Taking both obviously seems to be the better choice, but there is a catch: a superbeing—God, if you like—has the power of

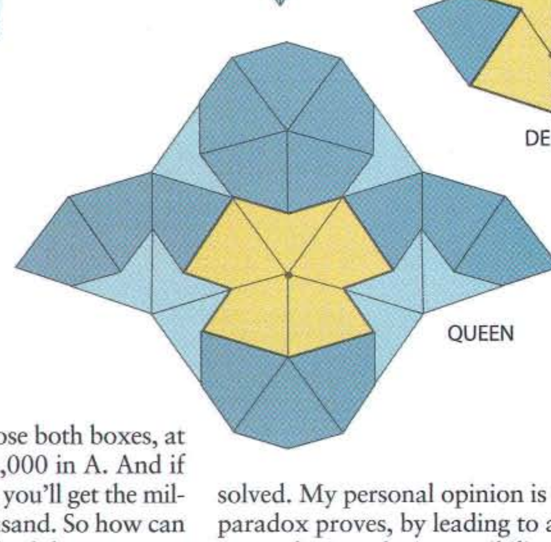


$$\phi = \frac{1+\sqrt{5}}{2}$$









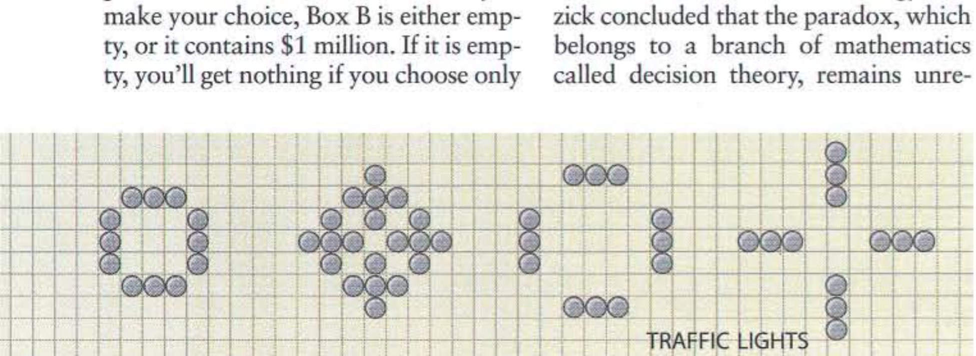
IAN WORPOLE

PENROSE TILES can be constructed by dividing a rhombus into a "kite" and a "dart" such that the ratio of their diagonals is phi (ϕ), the golden ratio (*above*). Arranging five of the darts around a vertex creates a star. Placing 10 kites around the star and extending the tiling symmetrically generate the infinite star pattern (*right*). Other tilings around a vertex include the deuce, jack and queen, which can also generate infinite patterns of kites and darts (*below right*).

knowing in advance how you will choose. If he predicts that out of greed you will take both boxes, he leaves B empty, and you will get only the $1,000 in A. But if he predicts you will take only Box B, he puts $1 million in it. You have watched this game played many times with others, and in every case when the player chose both boxes, he or she found that B was empty. And every time a player chose only Box B, he or she became a millionaire.

How should you choose? The pragmatic argument is that because of the previous games you have witnessed, you can assume that the superbeing does indeed have the power to make accurate predictions. You should therefore take only Box B to guarantee that you will get the $1 million. But wait! The superbeing makes his prediction *before* you play the game and has no power to alter it. At the moment you make your choice, Box B is either empty, or it contains $1 million. If it is empty, you'll get nothing if you choose only Box B. But if you choose both boxes, at least you'll get the $1,000 in A. And if B contains $1 million, you'll get the million plus another thousand. So how can you lose by choosing both boxes?

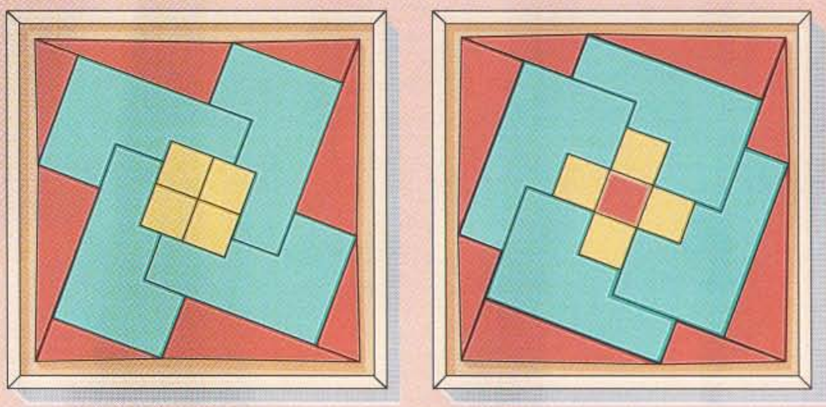
Each argument seems unassailable. Yet both cannot be the best strategy. Nozick concluded that the paradox, which belongs to a branch of mathematics called decision theory, remains unresolved. My personal opinion is that the paradox proves, by leading to a logical contradiction, the impossibility of a superbeing's ability to predict decisions. I wrote about the paradox in the July 1973 column and received so many letters afterward that I packed them into a carton and personally delivered them to Nozick. He analyzed the letters in a guest column in the March 1974 issue.



Magic squares have long been a popular part of recreational math. What makes these squares magical is the arrangement of numbers inside them: the numbers in every column, row and diagonal add up to the same sum. The numbers in the magic square are usually required to be distinct and run in

The Vanishing Area Paradox

Consider the figures shown below. Each pattern is made with the same 16 pieces: four large right triangles, four small right triangles, four eight-sided pieces and four small squares. In the pattern on the left, the pieces fit together snugly, but the pattern on the right has a square hole in its center! Where did this extra bit of area come from? And why does it vanish in the pattern on the left?



The secret to this paradox—which I devised for the "Mathematical Games" column in the May 1961 issue of *Scientific American*—will be revealed in the Letters to the Editors section of next month's issue. Impatient readers can find the answer at www.sciam.com on the World Wide Web. —*M.G.*

IAN WORPOLE

consecutive order, starting with one. There exists only one order-3 magic square, which arranges the digits one through nine in a three-by-three grid. (Variations made by rotating or reflecting the square are considered trivial.) In contrast, there are 880 order-4 magic squares, and the number of arrangements increases rapidly for higher orders.

Surprisingly, this is not the case with magic hexagons. In 1963 I received in the mail an order-3 magic hexagon devised by Clifford W. Adams, a retired clerk for the Reading Railroad. I sent the magic hexagon to Charles W. Trigg, a mathematician at Los Angeles City College, who proved that this elegant pattern was the only possible order-3 magic hexagon—and that no magic hexagons of any other size are possible!

What if the numbers in a magic square are not required to run in consecutive order? If the only requirement is that the numbers be distinct, a wide variety of order-3 magic squares can be constructed. For example, there is an infinite number of such squares that contain distinct prime numbers. Can an order-3 magic square be made with nine distinct square numbers? Two years ago in an article in *Quantum*, I offered $100 for such a pattern. So far no one has come forward with a "square of squares"—but no one has proved its impossibility either. If it exists, its numbers would be huge, perhaps beyond the reach of today's fastest supercomputers. Such a magic square would probably not have any practical use. Why then are mathematicians trying to find it? Because it might be there.

The Amazing Dr. Matrix

Every year or so during my tenure at *Scientific American*, I would devote a column to an imaginary interview with a numerologist I called Dr. Irving Joshua Matrix (note the "666" provided by the number of letters in his first, middle and last names). The good doctor would expound on the unusual properties of numbers and on bizarre forms of wordplay. Many readers thought Dr. Matrix and his beautiful, half-Japanese daughter, Iva Toshiyori, were real. I recall a letter from a puzzled Japanese reader who told me that Toshiyori was a most peculiar surname in Japan. I had taken it from a map of Tokyo. My informant said that in Japanese the word means "street of old men."

I regret that I never asked Dr. Matrix for his opinion on the preposterous 1997 best-seller *The Bible Code*, which claims to find predictions of the future in the arrangement of Hebrew letters in the Old Testament. The book employs a cipher system that would have made Dr. Matrix proud. By selectively applying this system to certain blocks of text, inquisitive readers can find hidden predictions not only in the Old Testament but also in the New Testament, the Koran, the *Wall Street Journal*—and even in the pages of *The Bible Code* itself.

The last time I heard from Dr. Matrix, he was in Hong Kong, investigating the accidental appearance of π in well-known works of fiction. He cited, for example, the following sentence fragment in chapter nine of book two of H. G. Wells's *The War of the Worlds:* "For a time I stood regarding..." The letters in the words give π to six digits! SA

The Author

MARTIN GARDNER wrote the "Mathematical Games" column for *Scientific American* from 1956 to 1981 and continued to contribute columns on an occasional basis for several years afterward. These columns are collected in 15 books, ending with *The Last Recreations* (Springer-Verlag, 1997). He is also the author of *The Annotated Alice, The Whys of a Philosophical Scrivener, The Ambidextrous Universe, Relativity Simply Explained* and *The Flight of Peter Fromm*, the last a novel. His more than 70 other books are about science, mathematics, philosophy, literature and his principal hobby, conjuring.

Further Reading

Recreations in the Theory of Numbers. Albert H. Beiler. Dover Publications, 1964.

Mathematics: Problem Solving through Recreational Mathematics. Bonnie Averbach and Orin Chein. W. H. Freeman and Company, 1986.

Mathematical Recreations and Essays. 13th edition. W. W. Rouse Ball and H.S.M. Coxeter. Dover Publications, 1987.

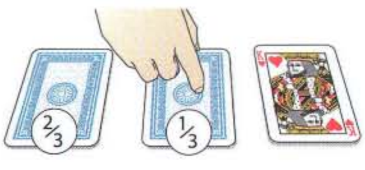
Penguin Edition of Curious and Interesting Geometry. David Wells. Penguin, 1991.

Mazes of the Mind. Clifford Pickover. St. Martin's Press, 1992.

Answers to the Four Gardner Puzzles

(The puzzles are on page 49.)

1. Most people guess that the probability has risen from $1/3$ to $1/2$. After all, only two cards are face down, and one must be the ace. Actually, the probability remains $1/3$. The probability that you *didn't* pick the ace remains $2/3$, but Jones has eliminated some of the uncertainty by showing that one of the two unpicked cards is not the ace. So there is a $2/3$ probability that the other unpicked card is the ace. If Jones gives you the option to change your bet to that card, you should take it (unless he's slipping cards up his sleeve, of course).



I introduced this problem in my October 1959 column in a slightly different form—instead of three cards, the problem involved three prisoners, one of whom had been pardoned by the governor. In 1990 Marilyn vos Savant, the author of a popular column in *Parade* magazine, presented still another version of the same problem, involving three doors and a car behind one of them. She gave the correct answer but received thousands of angry letters—many from mathematicians—accusing her of ignorance of probability theory! The fracas generated a front-page story in the *New York Times*.

2. The sum is 111. The trick always works because the matrix of numbers is nothing more than an old-fashioned addition table (*below*). The table is generated by two sets of numbers: (3, 1, 5, 2, 4, 0) and (25, 31, 13, 1, 7, 19). Each number in the matrix is the sum of a pair of numbers in the two sets. When you choose the six circled numbers, you are selecting six pairs that together include all 12 of the generating numbers. So the sum of the circled numbers is always equal to the sum of the 12 generating numbers. These special magic squares were the subject of my January 1957 column.

	3	1	5	2	4	0
25	28	~~26~~	~~30~~	~~27~~	~~29~~	~~25~~
31	~~34~~	~~32~~	~~36~~	~~33~~	35	~~31~~
13	~~16~~	~~14~~	18	~~15~~	~~17~~	~~13~~
1	~~4~~	~~2~~	~~6~~	3	~~5~~	~~1~~
7	~~10~~	8	~~12~~	~~9~~	~~11~~	~~7~~
19	~~22~~	~~20~~	~~24~~	~~21~~	~~23~~	19

3. Each chain of words ends on "God." This answer may seem providential, but it is actually the result of the Kruskal Count, a mathematical principle first noted by mathematician Martin Kruskal in the 1970s. When the total number of words in a text is significantly greater than the number of letters in the longest word, it is likely that any two arbitrarily started word chains will intersect at a keyword. After that point, of course, the chains become identical. As the text lengthens, the likelihood of intersection increases.

I discussed Kruskal's principle in my February 1978 column. Mathematician John Allen Paulos applies the principle to word chains in his upcoming book *Once upon a Number*.

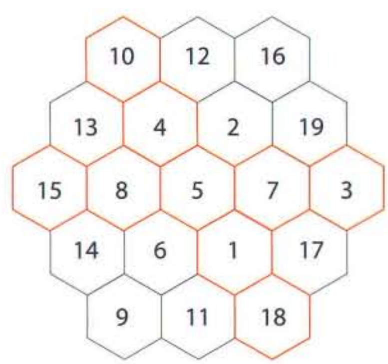
4. For simplicity's sake, imagine a deck of only 10 cards, with the black and red cards alternating like so: BRBRBRBRBR. Cutting this deck in half will produce two five-card decks: BRBRB and RBRBR. At the start of the shuffle, the bottom card of one deck is black, and the bottom card of the other deck is red. If the red card hits the table first, the bottom cards of both decks will then be black, so the next card to fall will create a black-red pair on the table. And if the black card drops first, the bottom cards of both decks will be red, so the next card to fall will create a red-black pair. After the first two cards drop—no matter which deck they came from—the situation will be the same as it was in the beginning: the bottom cards of the decks will be different colors. The process then repeats, guaranteeing a black and red card in each successive pair, even if some of the cards stick together (*below*).

THOROUGH SHUFFLE

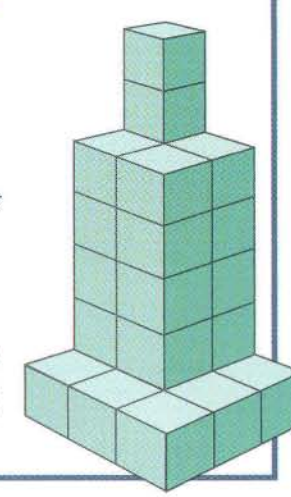
STICKY SHUFFLE

OR

This phenomenon is known as the Gilbreath principle after its discoverer, Norman Gilbreath, a California magician. I first explained it in my column in August 1960 and discussed it again in July 1972. Magicians have invented more than 100 card tricks based on this principle and its generalizations. —*M.G.*



MAGIC HEXAGON
has a unique property: every straight row of cells adds up to 38.



SKYSCRAPER
cannot be built from Soma pieces.
(The puzzle is on page 51.)

IAN WORPOLE

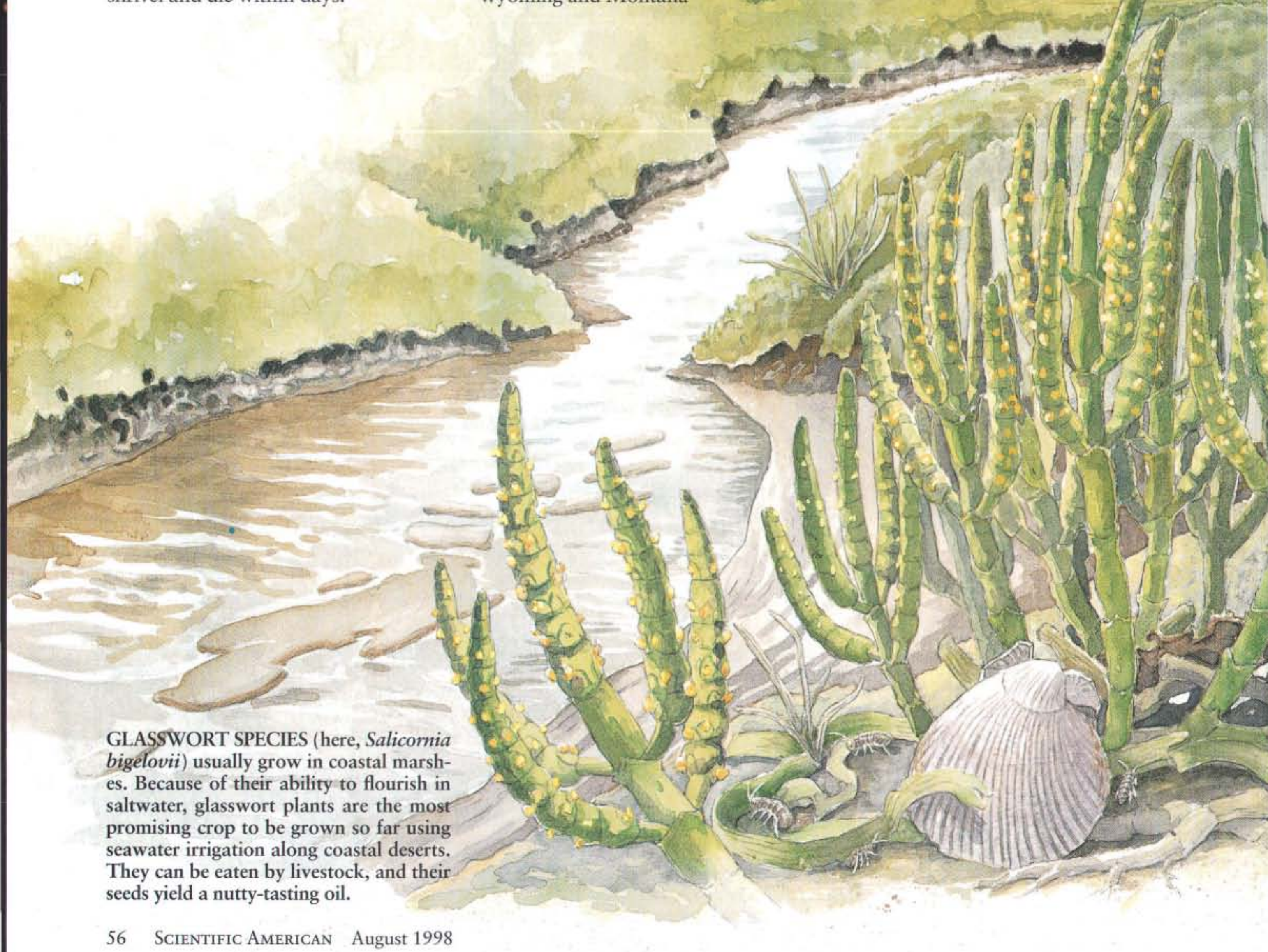
Irrigating Crops with Seawater

As the world's population grows and freshwater stores become more precious, researchers are looking to the sea for the water to irrigate selected crops

by Edward P. Glenn, J. Jed Brown and James W. O'Leary

Earth may be the Ocean Planet, but most terrestrial creatures—including humans—depend for food on plants irrigated by freshwater from rainfall, rivers, lakes, springs and streams. None of the top five plants eaten by people—wheat, corn, rice, potatoes and soybeans—can tolerate salt: expose them to seawater, and they droop, shrivel and die within days.

One of the most urgent global problems is finding enough water and land to support the world's food needs. The United Nations Food and Agriculture Organization estimates that an additional 200 million hectares (494.2 million acres) of new cropland—an area the size of Arizona, New Mexico, Utah, Colorado, Idaho, Wyoming and Montana



GLASSWORT SPECIES (here, *Salicornia bigelovii*) usually grow in coastal marshes. Because of their ability to flourish in saltwater, glasswort plants are the most promising crop to be grown so far using seawater irrigation along coastal deserts. They can be eaten by livestock, and their seeds yield a nutty-tasting oil.

RICHARD JONES

combined—will be needed over the next 30 years just to feed the burgeoning populations of the tropics and subtropics. Yet only 93 million hectares are available in these nations for farms to expand—and much of that land is forested and should be preserved. Clearly, we need alternative sources of water and land on which to grow crops.

With help from our colleagues, we have tested the feasibility of seawater agriculture and have found that it works well in the sandy soils of desert environments. Seawater agriculture is defined as growing salt-tolerant crops on land using water pumped from the ocean for irrigation. There is no shortage of seawater: 97 percent of the water on earth is in the oceans. Desert land is also plentiful: 43 percent of the earth's total land surface is arid or semiarid, but only a small fraction is close enough to the sea to make seawater farming feasible. We estimate that 15 percent of undeveloped land in the world's coastal and inland salt deserts could be suitable for growing crops using saltwater agriculture. This amounts to 130 million hectares of new cropland that could be brought into human or animal food production—without cutting down forests or diverting more scarce freshwater for use in agriculture.

Seawater agriculture is an old idea that was first taken seriously after World War II. In 1949 ecologist Hugo Boyko and horticulturalist Elisabeth Boyko went to the Red Sea town of Eilat during the formation of the state of Israel to create landscaping that would attract settlers. Lacking freshwater, the Boykos used a brackish well and seawater pumped directly from the ocean and showed that many plants would grow beyond their normal salinity limits in sandy soil [see "Salt-Water Agriculture," by Hugo Boyko; SCIENTIFIC AMERICAN, March 1967]. Although many of the Boykos' ideas of how plants tolerate salts have not stood the test of time, their work stimulated widespread interest, including our own, in extending the salinity constraints of traditional irrigated agriculture.

Seawater agriculture must fulfill two requirements to be cost-effective. First, it must produce useful crops at yields high enough to justify the expense of pumping irrigation water from the sea. Second, researchers must develop agronomic techniques for growing seawater-irrigated crops in a sustainable manner—one that does not damage the environment. Clearing these hurdles has proved a daunting task, but we have had some success.

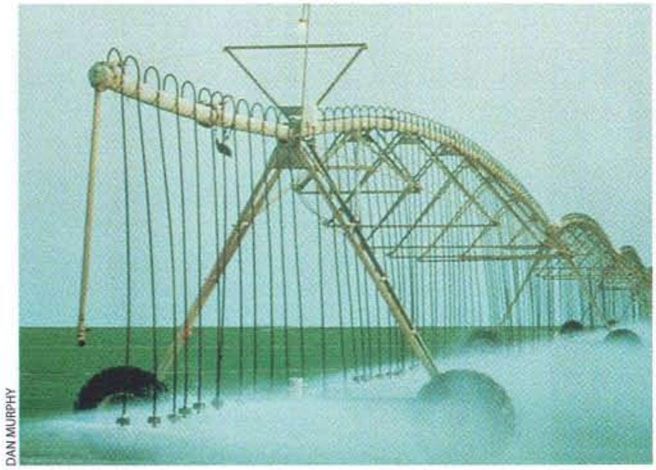
Salty Crops

The development of seawater agriculture has taken two directions. Some investigators have attempted to breed salt tolerance into conventional crops, such as barley and wheat. For example, Emanuel Epstein's research team at the University of California at Davis showed as early as 1979 that strains of barley propagated for generations in the presence of low levels of salt could produce small amounts of grain when irrigated by comparatively saltier seawater. Unfortunately, subsequent efforts to increase the salt tolerance of conventional crops through selective breeding and genetic engineering—in which genes for salt tolerance were added directly to the plants—have not produced good candidates for seawater irrigation. The upper salinity limit for the long-term irrigation of even the most salt-tolerant crops, such as the date palm, is still less than five parts per 1,000 (ppt)—less than 15 percent of the salt content of seawater. Normal seawater is 35 ppt salt, but in waters along coastal deserts, such as the Red Sea, the northern Gulf of California (between the western coast of Sonora in Mexico and Baja California) and the Persian Gulf, it is usually closer to 40 ppt. (Sodium chloride, or table salt, is the most prevalent salt in seawater and the one that is most harmful to plant growth.)

Our approach has been to domesticate wild, salt-tolerant plants, called halophytes, for use as food, forage and oilseed crops. We reasoned that changing the basic physiology of a traditional crop plant from salt-sensitive to salt-tolerant would be difficult and that it might be more feasible to domesticate a wild, salt-tolerant plant. After all, our modern crops started out as wild plants. Indeed, some halophytes—such as grain from the saltgrass *Distichlis palmeri* (Palmer's grass)—were eaten for generations by native peoples, including the Cocopah, who live where the Colorado River empties into the Gulf of California.

We began our seawater agriculture efforts by collecting several hundred halophytes from throughout the world and screening them for salt tolerance and nutritional content in the laboratory. There are between 2,000 and 3,000 species of halophytes, from grasses and

UNIVERSITY OF ARIZONA



DAN MURPHY

SEAWATER AGRICULTURE can require different agronomic techniques than freshwater agriculture. To grow saltbush, or *Atriplex*—a salt-tolerant plant that can be used to feed livestock—seawater farmers must flood their fields frequently (*left*). In addition, irrigation booms (*center*) must be lined with plastic piping to protect them from rusting when in contact with the salty water. But some techniques can remain the same: standard combines are used to harvest *Salicornia* seeds (*right*), for example.

shrubs to trees such as mangroves; they occupy a wide range of habitats—from wet, seacoast marshes to dry, inland saline deserts. In collaboration with Dov Pasternak's research team at Ben Gurion University of the Negev in Israel and ethnobotanists Richard S. Felger and Nicholas P. Yensen—who were then at the University of Arizona—we found roughly a dozen halophytes that showed sufficient promise to be grown under agronomic conditions in field trials.

In 1978 we began trials of the most promising plants in the coastal desert at Puerto Peñasco, on the western coast of Mexico. We irrigated the plants daily by flooding the fields with high-saline (40 ppt) seawater from the Gulf of California. Because the rainfall at Puerto Peñasco averages only 90 millimeters a year—and we flooded our plots with an annual total depth of 20 meters or more of seawater—we were certain the plants were growing almost solely on seawater. (We calculate rainfall and irrigation according to the depth in meters that falls on the fields rather than in cubic meters, which is a measure of volume.)

Although the yields varied among species, the most productive halophytes produced between one and two kilograms per square meter of dry biomass—roughly the yield of alfalfa grown using freshwater irrigation. Some of the most productive and salt-tolerant halophytes were shrubby species of *Salicornia* (glasswort), *Suaeda* (sea blite) and *Atriplex* (saltbush) from the family Chenopodiaceae, which contains about 20 percent of all halophyte species. Salt grasses such as *Distichlis* and viny, succulent-leaved ground covers such as *Batis* were also highly productive. (These plants are not Chenopodiaceae, though; they are members of the Poaceae and Batidaceae families, respectively.)

But to fulfill the first cost-effectiveness requirement for seawater agriculture, we had to show that halophytes could replace conventional crops for a specific use. Accordingly, we tested whether halophytes could be used to feed livestock. Finding enough forage for cattle, sheep and goat herds is one of the most challenging agricultural problems in the world's drylands, 46 percent of which have been degraded through overgrazing, according to the U.N. Environment Program. Many halophytes have high levels of protein and digestible carbohydrates. Unfortunately, the plants also contain large amounts of salt; accumulating salt is one of the ways they adjust to a saline environment [*see illustration on page 60*]. Because salt has no calories yet takes up space, the high salt content of halophytes dilutes their nutritional value. The high salinity of halophytes also limits the amount an animal can eat. In open grazing situations, halophytes are usually considered "reserve-browse plants," to which animals turn only when more palatable plants are gone.

Our strategy was to incorporate halophytes as part of a mixed diet for livestock, replacing conventional hay forage with halophytes to make up between 30 and 50 percent of the total food intake of sheep and goats. (These percentages are the typical forage levels used in fattening animals for slaughter.) We found that animals fed diets containing *Salicornia, Suaeda* and *Atriplex* gained as much weight as those whose diets included hay. Moreover, the quality of the test animals' meat was unaffected by their eating a diet rich in halophytes. Contrary to our initial fears, the animals had no aversion to eating halophytes in mixed diets; they actually seemed to be attracted by the salty taste. But the animals that ate a halophyte-rich diet drank more water than those that ate hay, to compensate for the extra salt intake. In addition, the feed conversion ratio of the test animals (the amount of meat they produced per kilogram of feed) was 10 percent lower than that of animals eating a traditional diet.

Farming for Oil

The most promising halophyte we have found thus far is *Salicornia bigelovii.* It is a leafless, succulent, annual salt-marsh plant that colonizes new areas of mud flat through prolific seed production. The seeds contain high levels of oil (30 percent) and protein (35 percent), much like soybeans and other oilseed crops, and the salt content is less than 3 percent. The oil is highly polyunsaturated and similar to safflower oil in fatty-acid composition. It can be extracted from the seed and refined using conventional oilseed equipment; it is also edible, with a pleasant, nutlike taste


DAN MURPHY

and a texture similar to olive oil. A small drawback is that the seed contains saponins, bitter compounds that make the raw seeds inedible. These do not contaminate the oil, but they can remain in the meal after oil extraction. The saponins thus restrict the amount of meal that can be used in chicken diets, but feeding trials have shown that *Salicornia* seed meal can replace conventional seed meals at the levels normally used as a protein supplement in livestock diets. Hence, every part of the plant is usable.

We have participated in building several prototype *Salicornia* farms of up to 250 hectares in Mexico, the United Arab Emirates, Saudi Arabia and India. During six years of field trials in Mexico, *Salicornia* produced an average annual crop of 1.7 kilograms per square meter of total biomass and 0.2 kilogram per square meter of oilseed. These yields equal or exceed the yields of soybeans and other oilseeds grown using freshwater irrigation. We have also shown that normal farm and irrigation equipment can be modified so that it is protected from salt damage from the seawater. Although the irrigation strategies for handling seawater are different from those used for freshwater crops, we have not encountered any insurmountable engineering problems in scaling up from field tests to prototype farms.

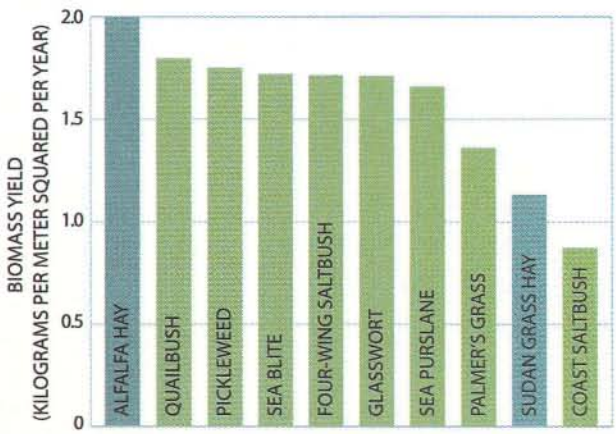
Normally, crops are irrigated only when the soil dries to about 50 percent of its field capacity, the amount of water it is capable of holding. In addition, in freshwater irrigation, farmers add only enough water to replace what the plants have used. In contrast, seawater irrigation requires copious and frequent—even daily—irrigation to prevent salt from building up in the root zone to a level that inhibits growth.

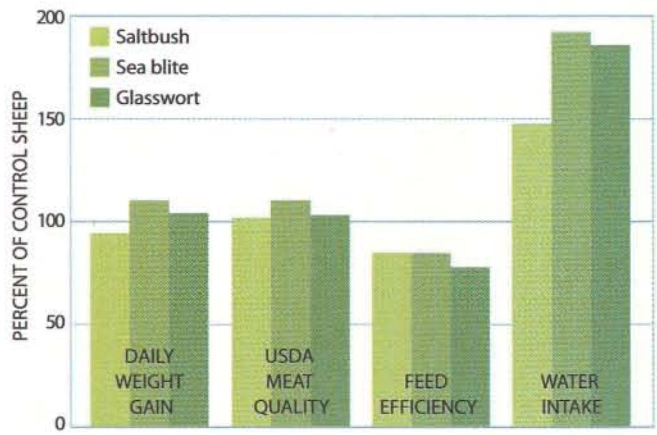
Our first field trials used far more water (20 meters a year) than could be applied economically, so in 1992 we began experiments to determine the minimum amount of seawater irrigation needed to produce a good yield. We grew test plantings of *Salicornia* in soil boxes buried within open, irrigated plots of the same crop during two years of field trials. The boxes, called lysimeters, had bottom drains that conveyed excess water to several collection points outside the plots, allowing us to measure the volume and salinity of the drain water. Using them, we calculated the water and salt balances required for a seawater-irrigated crop for the first time. We found that the amount of biomass a seawater-irrigated crop yields depends on the amount of seawater used. Although *Salicornia* can thrive when the salinity of the water bathing its roots exceeds 100 ppt—roughly three times the normal saltiness of the ocean—it needs approximately 35 percent more irrigation when grown using seawater than conventional crops grown using freshwater. *Salicornia* requires this extra water because as it selectively absorbs water from the seawater, it quickly renders the remaining seawater too salty for use.

Making It Pay

Can seawater agriculture be economical? The greatest expense in irrigated agriculture is in pumping the water. The pumping costs are directly proportional to the amount of water pumped and the height to which it is lifted. Although halophytes require more water than conventional crops, seawater farms near sea level require less water lifting than conventional farms, which often lift water from wells deeper than 100 meters. Because pumping seawater at sea level is cheaper than pumping freshwater from wells, seawater agriculture should be cost-effective in desert regions—even though its yields are smaller than traditional, freshwater agriculture.

Seawater irrigation does not require special equipment. The large test farms we have helped build have used either flood irrigation of large basins or moving-boom sprinkler irrigation. Moving booms are used in many types of crop





SLIM FILMS

YIELDS of salt-tolerant crops grown using seawater agriculture are comparable to those of two freshwater-irrigated plants often used for livestock forage: alfalfa hay and Sudan grass hay (*left, blue bars*). Sheep raised on a diet supplemented with salt-tolerant plants such as saltbush, sea blite and glasswort gain at least as much weight and yield meat of the same quality as control sheep fed conventional grass hay, although they convert less of the feed to meat and must drink almost twice as much water (*right*).

production. For seawater use, a plastic pipe is inserted in the boom so the seawater does not contact metal. *Salicornia* seeds have also been successfully harvested using ordinary combines set to maximize retention of the very small seeds, which are only roughly one milligram in weight.

Yet *Salicornia,* our top success story so far, is not a perfect crop. The plants tend to lodge (lie flat in the field) as harvest approaches, and the seeds may shatter (release before harvest). In addition, seed recoveries are only about 75 percent for *Salicornia,* compared with greater than 90 percent for most crops. Further, to support high seed yields *Salicornia* must grow for approximately 100 days at cool temperatures before flowering. Currently production of this crop is restricted to the subtropics, which have cool winters and hot summers; however, some of the largest areas of coastal desert in the world are in the comparatively hotter tropics.

The second cost-effectiveness requirement of seawater agriculture is sustainability over the long term. But sustainability is not a problem limited to irrigation using seawater: in fact, many irrigation projects that use freshwater cannot pass the sustainability test. In arid regions, freshwater irrigation is often practiced in inland basins with restricted drainage, resulting in the buildup of salt in the water tables underneath the fields. Between 20 and 24 percent of the world's freshwater-irrigated lands suffer from salt and water buildup in the root zone. When the problem becomes severe, farmers must install expensive subsurface drainage systems; disposing of the collected drain water creates additional problems. In California's San Joaquin Valley, for example, wastewater that had drained into a wetland caused death and deformity in waterfowl because of the toxic effects of selenium, an element that typically occurs in many western U.S. soils but had built up to high concentrations in the drain water.

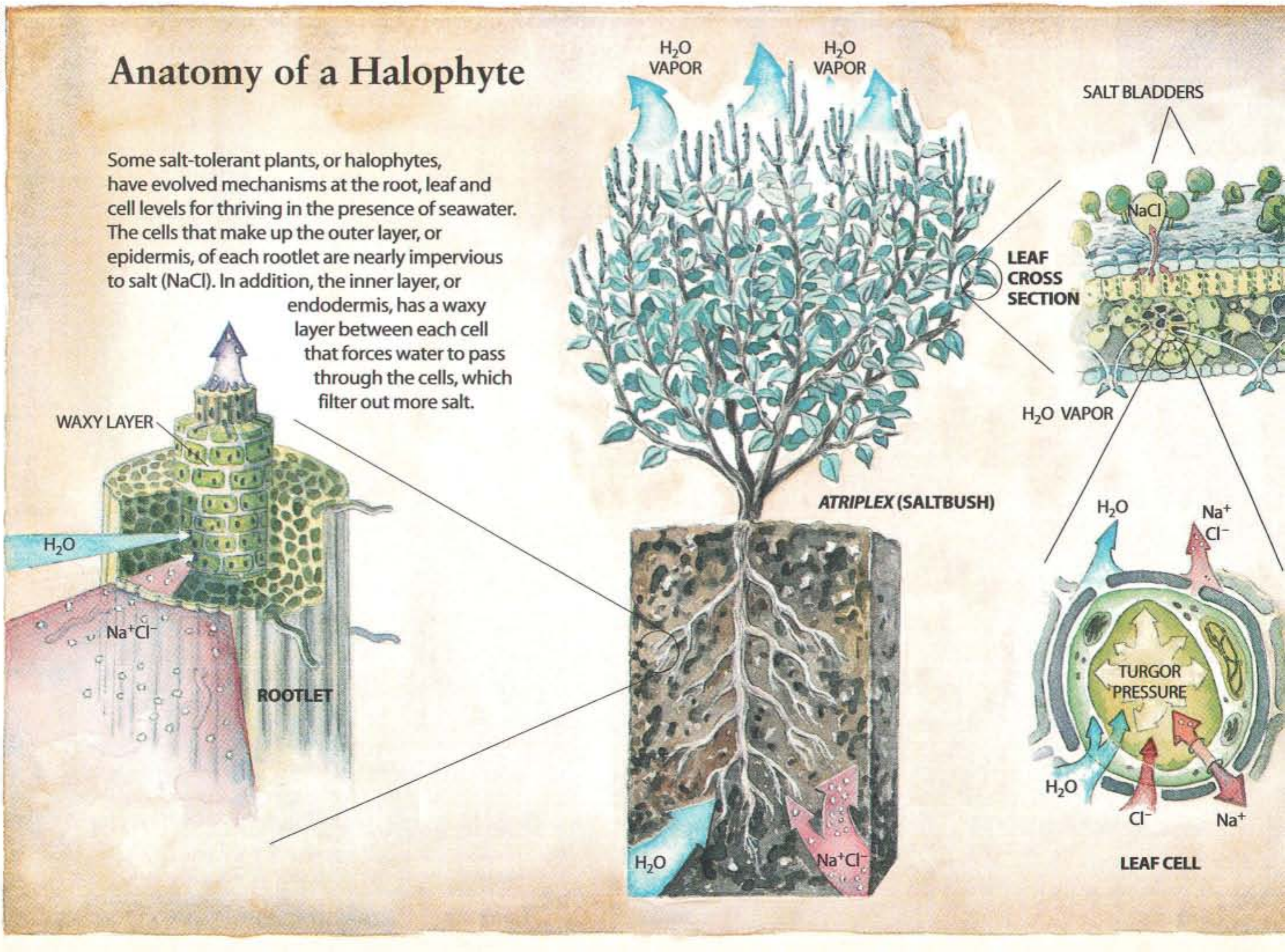
Seawater agriculture is not necessarily exempt from such problems, but it does offer some advantages. First, coastal desert farms on sandy soils generally have unimpeded drainage back to the sea. We have continuously irrigated the same fields with seawater for over 10 years with no buildup of water or salts in the root zone. Second, coastal and inland salt desert aquifers often already have elevated concentrations of salt and so should not be damaged by seawater. Third, the salt-affected soils that we propose for seawater agriculture are often barren—or nearly so—to start with, so installing a seawater farm may have far less effect on sensitive ecosystems than conventional agriculture does.

No farming activity is completely benign, however. Large-scale coastal shrimp farms, for example, have caused algal blooms and disease problems in rivers or bays that receive their nutrient-rich effluent [see "Shrimp Aquaculture and the Environment," by Claude E. Boyd and Jason W. Clay; SCIENTIFIC AMERICAN, June]. A similar problem can be anticipated from large-scale halophyte farms, caused by the large volume of high-salt drainage water containing unused fertilizer, which will ultimately be discharged back to the sea. On the other hand, seawater farms can also be part of a solution to this problem if shrimp-farm effluent is recycled onto a halophyte farm instead of dis-

RICHARD JONES

Anatomy of a Halophyte

Some salt-tolerant plants, or halophytes, have evolved mechanisms at the root, leaf and cell levels for thriving in the presence of seawater. The cells that make up the outer layer, or epidermis, of each rootlet are nearly impervious to salt (NaCl). In addition, the inner layer, or endodermis, has a waxy layer between each cell that forces water to pass through the cells, which filter out more salt.



LOCATIONS in coastal deserts and inland salt deserts (*green areas*) could be used for seawater agriculture—or agriculture using irrigation from salty underground aquifers—to grow a variety of salt-tolerant crops for food or animal forage.

charged directly to the sea: the halophyte crop would recover many of the nutrients in the effluent and reduce the volume. The first halophyte test farm we built in Mexico was installed to recycle shrimp-farm effluent, and further research linking marine aquaculture effluent with halophyte farms is under way.

Halophyte farms have also been proposed as a way to recycle the selenium-rich agricultural drain water generated in the San Joaquin Valley of California. Selenium is an essential nutrient at low levels but becomes toxic at high levels. Halophytes grown on drain water in the valley take up enough selenium to make them useful as animal-feed supplements but not enough to make them toxic.

Will seawater agriculture ever be practiced on a large scale? Our goal in the late 1970s was to establish the feasibility of seawater agriculture; we expected to see commercial farming within 10 years. Twenty years later seawater agriculture is still at the prototype stage of commercial development. Several companies have established halophyte test farms of *Salicornia* or *Atriplex* in California, Mexico, Saudi Arabia, Egypt, Pakistan and India; however, to our knowledge, none have entered large-scale production. Our research experience convinces us of the feasibility of seawater agriculture. Whether the world ultimately turns to this alternative will depend on future food needs, economics and the extent to which freshwater ecosystems are withheld from further agricultural development.

Some halophytes, such as this *Atriplex* (saltbush), have specialized cells called salt bladders on their leaves for storing excess NaCl. When full, the salt bladders burst, releasing salt in a silvery layer that reflects light and cools the leaf. Water vapor escapes through pores on the undersides of the leaves.

Cells inside each leaf are specially equipped to handle any salt that is absorbed by the plant. The central vacuole, or storage area, of each cell bears molecules that specifically import sodium ions (Na^+), and chloride ions (Cl^-) follow. The high concentration of Na^+ and Cl^- attracts water, maintaining the turgor pressure of the cell.

The Authors

EDWARD P. GLENN, J. JED BROWN and JAMES W. O'LEARY have a combined total of 45 years of experience studying the feasibility of seawater agriculture in desert environments. Glenn began his research career as a self-described "marine agronomist" in 1978 after receiving his Ph.D. from the University of Hawaii; he is now a professor in the department of soil, water and environmental science at the University of Arizona at Tucson. Brown received his Ph.D. from the University of Arizona's Wildlife and Fisheries Program in May. O'Leary is a professor in the University of Arizona's department of plant sciences. He received his Ph.D. from Duke University in 1963. Author of more than 60 publications on plant sciences, O'Leary served in 1990 on a National Research Council panel that examined the prospects of seawater agriculture for developing countries.

Further Reading

SALINE CULTURE OF CROPS: A GENETIC APPROACH. Emanuel Epstein et al. in *Science*, Vol. 210, pages 399–404; October 24, 1980.

SALINE AGRICULTURE: SALT TOLERANT PLANTS FOR DEVELOPING COUNTRIES. National Academy Press, 1990.

SALICORNIA BIGELOVII TORR.: AN OILSEED HALOPHYTE FOR SEAWATER IRRIGATION. E. P. Glenn, J. W. O'Leary, M. C. Watson, T. L. Thompson and R. O. Kuehl in *Science*, Vol. 251, pages 1065–1067; March 1, 1991.

TOWARDS THE RATIONAL USE OF HIGH SALINITY TOLERANT PLANTS. H. Lieth and A. A. Al Masoom. Series: *Tasks for Vegetation Science*, Vol. 28. Kluwer Academic Publishers, 1993.

HALOPHYTES. E. P. Glenn in *Encyclopedia of Environmental Biology*. Academic Press, 1995.

Microdiamonds

These tiny, enigmatic crystals hold promise both for industry and for the study of how diamond grows

by Rachael Trautman, Brendan J. Griffin and David Scharf

Diamond has been a highly valued mineral for 3,000 years. In ancient India the gemstone was thought to possess magical powers conferred by the gods. Medieval European nobles sometimes wore diamond rings into battle, thinking the jewelry would bring them courage and make them fearless.

More recently, diamond has proved to be the ideal material for a number of industrial uses. The mineral, an incredibly pure composition of more than 99 percent carbon, is the hardest substance known. It is capable of scratching almost anything, making it suitable for use in abrasives and in cutting, grinding and polishing tools. Diamond also has high thermal conductivity—more than three times that of copper—and is thus optimal for spreading and dissipating heat in electronic devices such as semiconductor lasers. Because most of these applications can be accomplished with tiny crystals, both scientific and technological interest have begun to focus on microdiamonds, samples that measure less than half a millimeter in any dimension.

Previously, microdiamonds were ignored, mainly because of their minimal size and the complex and expensive procedure required to mine the tiny particles. But as extraction techniques improve and industrial demand grows, microdiamonds are becoming an increasingly attractive source of the mineral.

Mysterious Origins

After decades of research, scientists now believe that commercial-size stones, or macrodiamonds, form in the earth's mantle and are transported to the surface by volcanoes. The origin of microdiamonds, however, remains a mystery. They may simply be younger crystals that had little time to grow before being brought to the earth's surface. Or they may have formed in an environment where carbon was limited, thereby stunting their growth. Some researchers believe, however, that microdiamonds are produced by different, though related, processes.

Microdiamonds are occasionally found by themselves, in the absence of bigger crystals. Often, though, they occur along with macrodiamonds. Surprisingly, microdiamonds have even been discovered in the same deposits with larger samples that have experienced resorption—a broad term encompassing any process, such as dissolution or corrosion, that reduces the size of a diamond, frequently resulting in the rounding of corners or edges. Because the ratio of surface area to volume is much higher for micro specimens than for macro sizes, scientists would have expected the resorption that reduced the size of the macrodiamonds to have consumed any small samples.

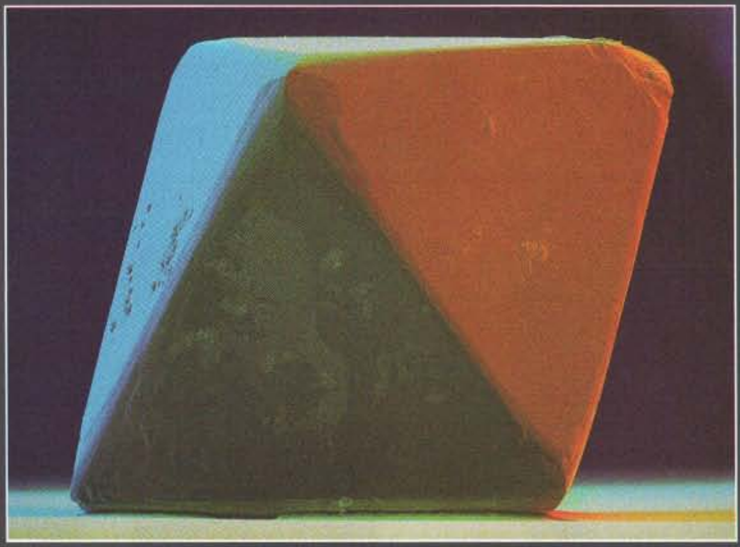
This paradoxical evidence suggests that microdiamonds might have origins distinct from those of macrodiamonds. One hypothesis asserts that some microdiamonds form within ascending magma, where distinct growth conditions and processes limit the size to which the carbon crystals can grow. But an alternative theory contends that the magma is merely the delivery mechanism and that microdiamonds, like macrodiamonds, are products of the mantle.

Complicating this issue, researchers have discovered microdiamonds that they believe were formed in the earth's crust from the collision of tectonic plates. And tiny diamonds have also been found in meteorites.

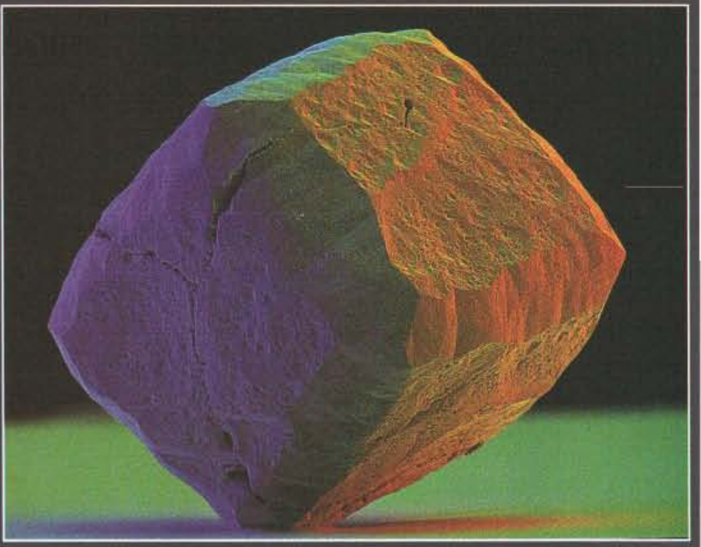
With continuing work, scientists may one day resolve how microdiamonds are related to their larger siblings. This knowledge will lead to a greater understanding of the carbon substance, perhaps resulting in more effective techniques for mining the mineral economically and for growing it synthetically.

ALL MICROGRAPHS BY DAVID SCHARF

MICRODIAMOND is smaller than half a millimeter (500 microns) in any dimension, corresponding to a mass of less than a hundredth of a carat. (In comparison, the famous Hope diamond has a mass of 45 carats.) This particular specimen is about 400 microns tall.



OCTAHEDRON



CUBE

A Variety of Shapes

Mineralogists believe that a diamond gets larger by adding new atoms to its surface, with the speed of growth influencing the shape of the crystal. If the temperature, pressure and oxygen content are favorable, then the growth rate—and final crystal form—will depend on how much carbon is available.

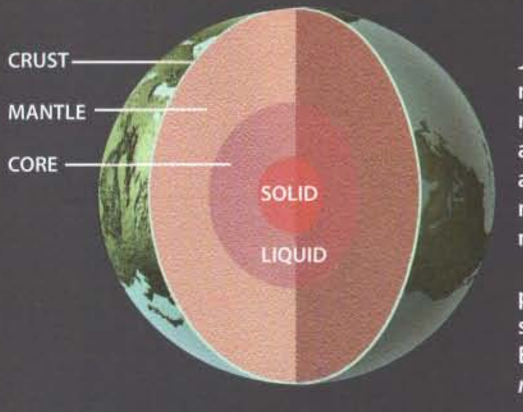
Microdiamonds occur as single crystals of octahedrons, dodecahedroids, cubes, macles (twinned, flat triangular plates) and irregular lumps. In addition, octahedral and cubic faces can sometimes develop concurrently in the same crystal, resulting in a cubo-octahedral form. Frequently, two or more microdiamonds will fuse during their growth, leading to unusual and often striking configurations such as aggregates of octahedrons, dodecahedroids or cubes; and amorphous clusters.

Octahedrons (*left*) are the most common growth form. In fact, in most microdiamond populations, they usually make up more than half the observed crystal shapes.

Less frequently found are cubes (*center*), which rarely have a crystalline or ordered atomic structure. Instead many cubic microdiamonds have a fibrous composition with radial growth perpendicular to the cubic faces. This distinctive form of growth, as indicated in the textured faces, is the result of supersaturation of carbon in the source medium, which leads to extremely rapid crystallization.

Aggregates (*right*) also imply an environment high in carbon, with growth occurring simultaneously at several neighboring points. In such proximity, the crystals may meld to form a single unit of complex intergrowth.

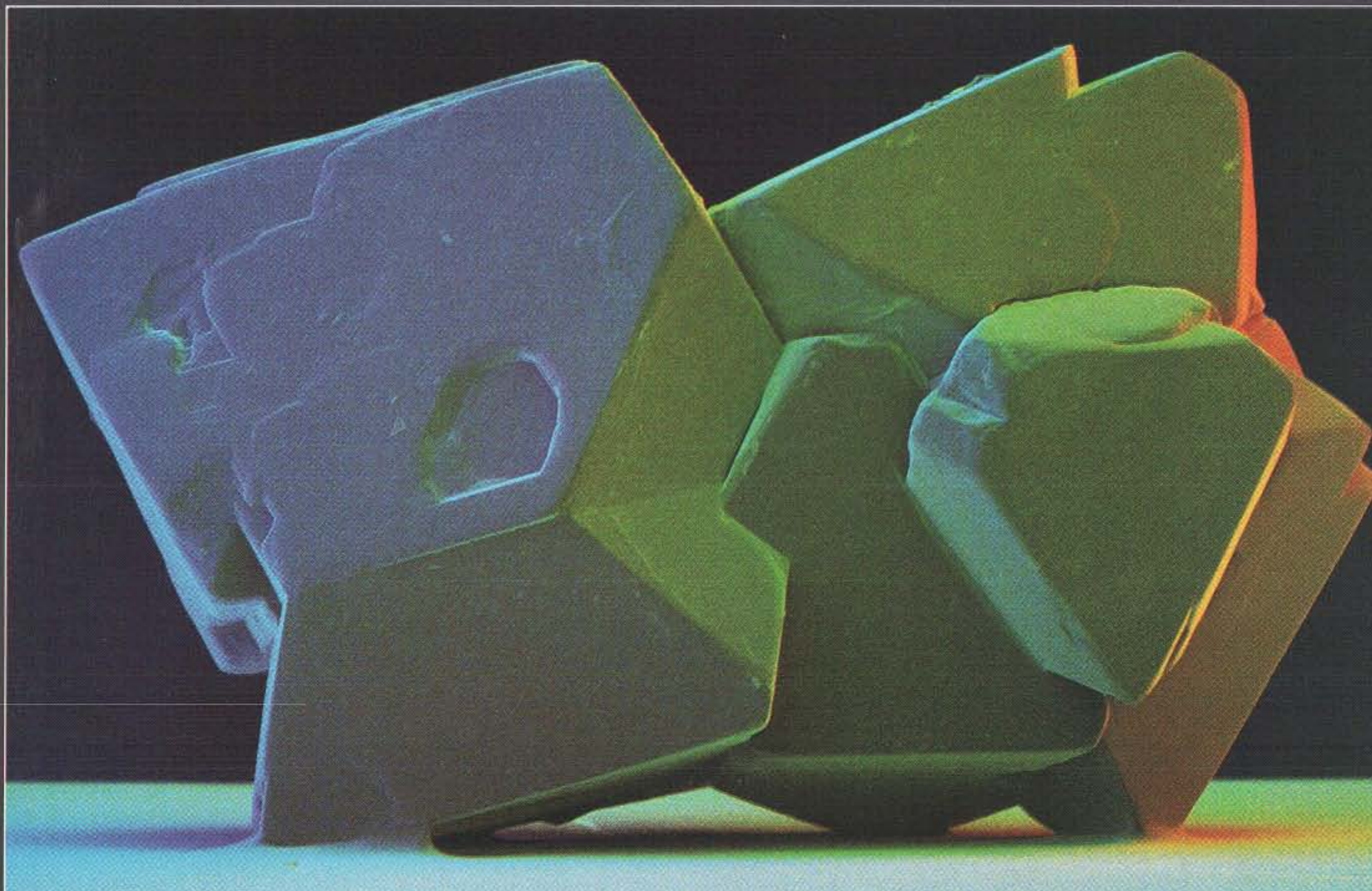
The Origin of Diamond



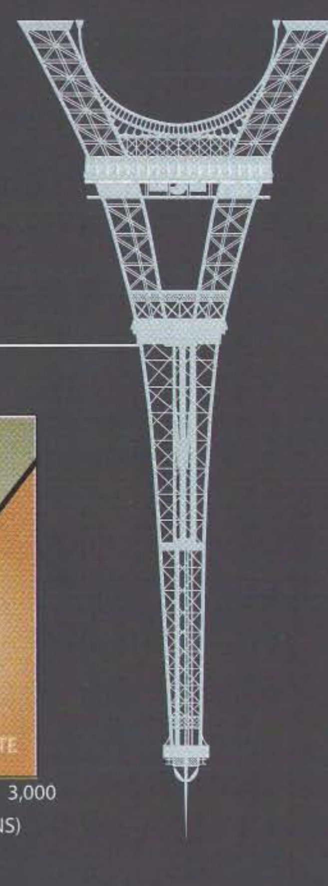
ILLUSTRATIONS BY LAURIE GRACE

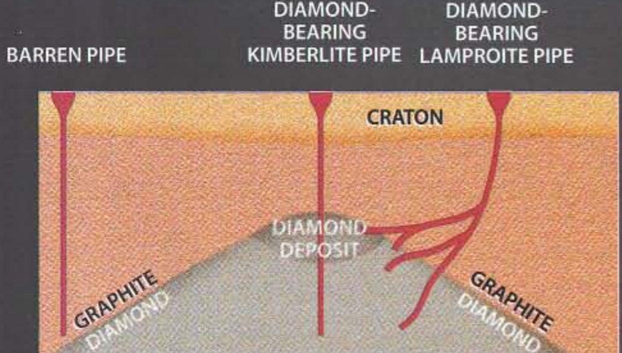
After much research, scientists have concluded that diamond forms in the earth's mantle (*left*), at depths of 150 to 200 kilometers (90 to 120 miles). The valuable mineral is then carried upward by kimberlites and lamproites, special types of igneous rock that began as molten material. At the earth's surface (*near right*), the rock is found as "pipes," or carrot-shaped deposits, predominantly within very old landmasses known as cratons. Such deposits may produce as little as one carat per 100 metric tons of rock mined. Interestingly, some researchers believe that microdiamonds—unlike macrodiamonds—do not form in the mantle but rather in the kimberlitic and lamproitic magma.

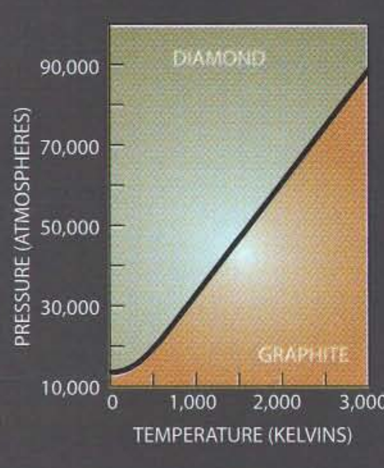
Whatever the case, tremendous pressure is required for diamond to form. At a temperature of 1,600 kelvins (2,400 degrees Fahrenheit), a pressure of about 50,000 atmospheres is needed (*center right*)—roughly equivalent to the pressure of an upside-down Eiffel Tower (all 7,000 metric tons of it) balanced on a 12-centimeter-square plate (*far right*). Without that extreme compression, the carbon may instead form graphite.



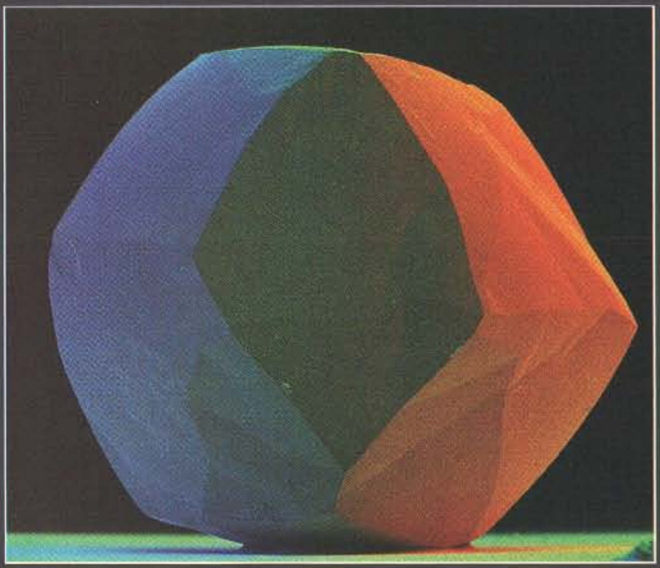
AGGREGATE



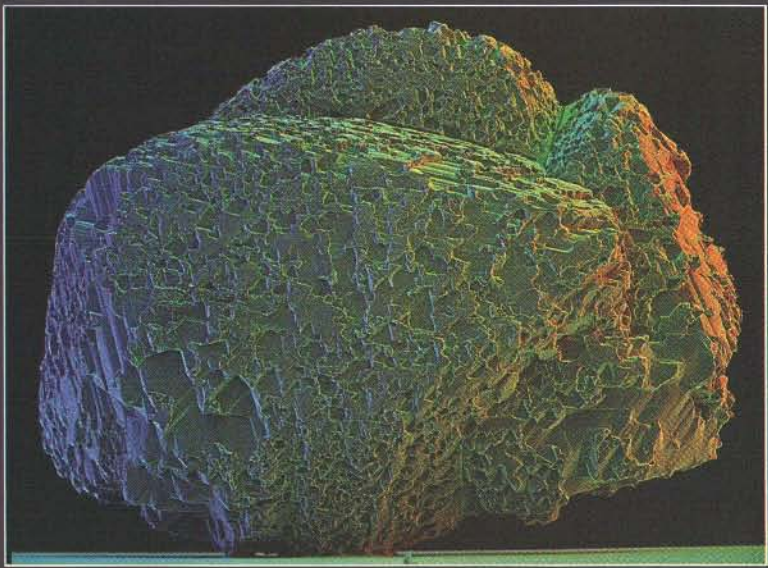




Microdiamonds



DODECAHEDROID formed by octahedral dissolution



OCTAHEDRAL AGGREGATE etched by corrosion

Changing Faces

Resorption processes typically alter the shape of a microdiamond. Dissolution, for example, converts plane crystal faces into curved or rounded surfaces. Specifically, microdiamonds do not grow as dodecahedroids (*left*); the 12-sided shape is instead the end product of octahedral dissolution. With sharp-edged octahedrons at one end of the spectrum and rounded dodecahedroids at the other, microdiamonds have exhibited a complete gradation between those two extremes. Because dodecahedroids are almost always found in microdiamond populations, dissolution is believed to be common.

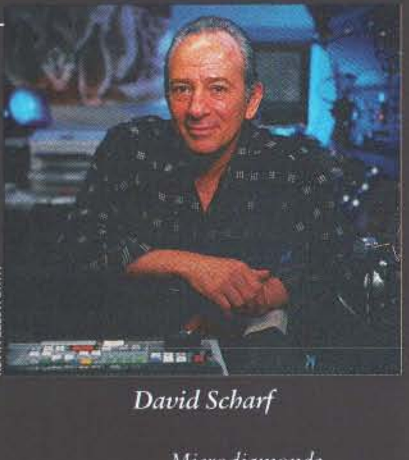
When microdiamonds are dissolved in the presence of a corrosive agent, strongly developed etching or unusually rough surfaces may develop. The erosion can often be extreme, producing diamond morphologies with almost no resemblance to their original forms, such as the transformation of the smooth, planar faces of an octahedral aggregate into etch-sculptured surfaces (*center*).

Sometimes the corrosion will result in faces with distinctive structures, such as those resembling steps. Particularly at the corners and edges, the etching can be severe enough to alter a crystal's overall shape (*right*).

"Seeing" the Microscopic World

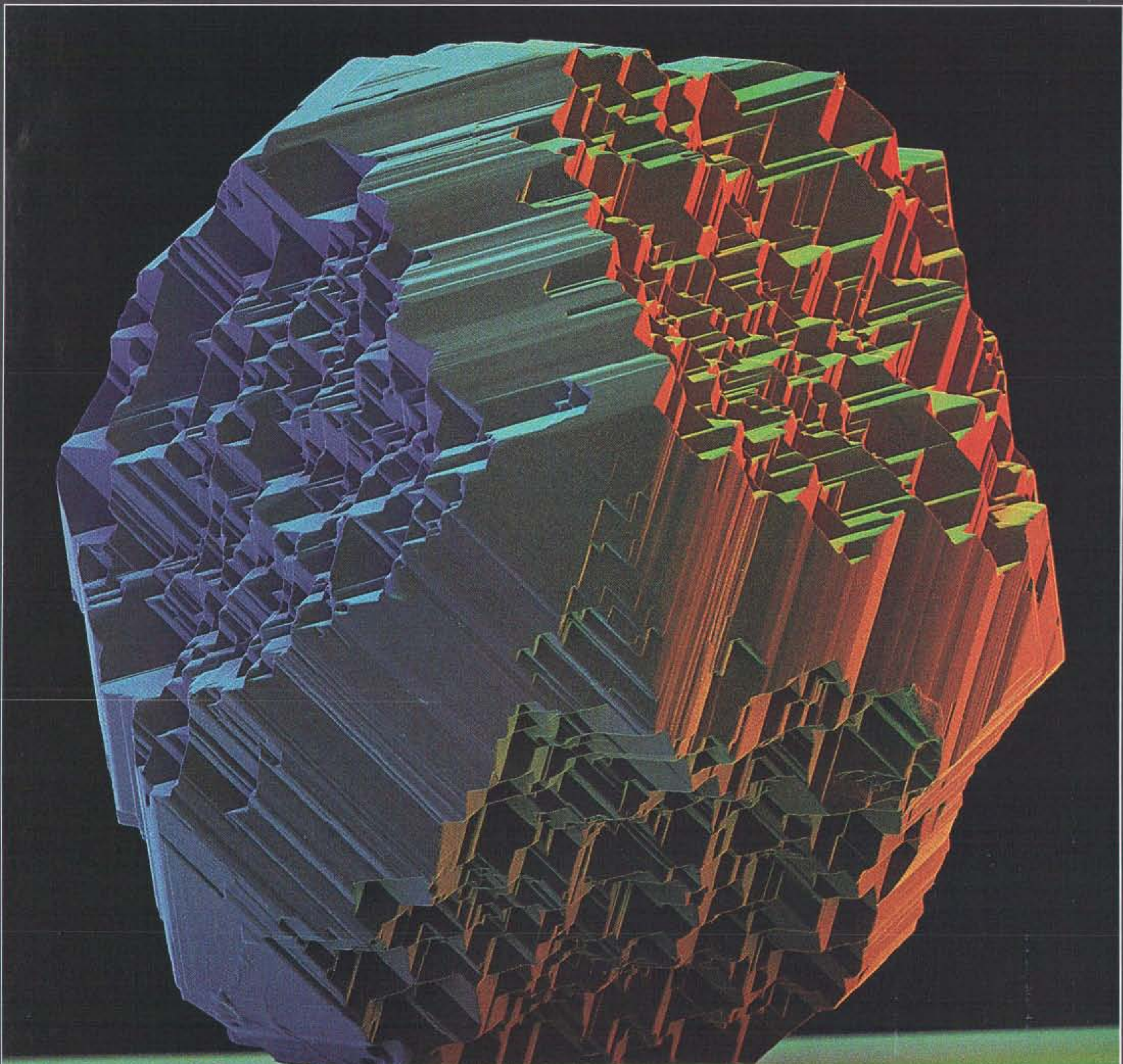
The microdiamond images presented here were recorded with a scanning electron microscope (SEM), by micrographer David Scharf (*right*). The SEM works by illuminating a subject with electrons rather than light. Magnetic lenses focus a beam of electrons that is scanned across the specimen in a raster pattern (similar to that used in television screens). When the beam hits the specimen, it knocks other (secondary) electrons from the surface of the object. These negatively charged particles are then attracted to the positive polarity of a detector, where they are amplified and converted into a video signal.

The microdiamonds appear to be opaque in the images because the crystals are not transparent to electrons. The color comes from a system developed by Scharf—the Multi-Detector Color Synthesizer (U.S. Patent No. 5,212,383)—which uses three electron detectors (instead of the usual one) spaced apart to capture three distinct illumination angles. A different (arbitrary) color is assigned to each detector to produce the various colors seen.



KEVIN ELLSWORTH

David Scharf



OCTAHEDRON etched into near cubic form

The Authors

RACHAEL TRAUTMAN, BRENDAN J. GRIFFIN and DAVID SCHARF met through their interest in microdiamonds and electron microscopy. Trautman is a postgraduate with the Center for Strategic Mineral Deposits in the department of geology and geophysics at the University of Western Australia, where she is studying the origin and genesis of microdiamonds. Griffin is a senior lecturer at the Center for Microscopy and Microanalysis at the University of Western Australia; his research interests include diamond genesis, mineral microanalysis and new techniques in scanning electron microscopy. Scharf is a professional scanning electron micrographer whose work has appeared in encyclopedias, museums and magazines such as *Scientific American.*

Further Reading

DIAMONDS. Second edition. Eric Bruton. NAG Press, London, 1978.

DIAMOND. Gordon Davies. Adam Hilger, Bristol, 1984.

A COMPARISON OF THE MICRODIAMONDS FROM KIMBERLITE AND LAMPROITE OF YAKUTIA AND AUSTRALIA. R. L. Trautman, B. J. Griffin, W. R. Taylor, Z. V. Spetsius, C. B. Smith and D. C. Lee in *Proceedings of the Sixth International Kimberlite Conference,* Vol. 2: *Russian Geology and Geophysics,* No. 38. Allerton Press, 1997.

THE NATURE OF DIAMONDS. Edited by George E. Harlow. Cambridge University Press, 1998.

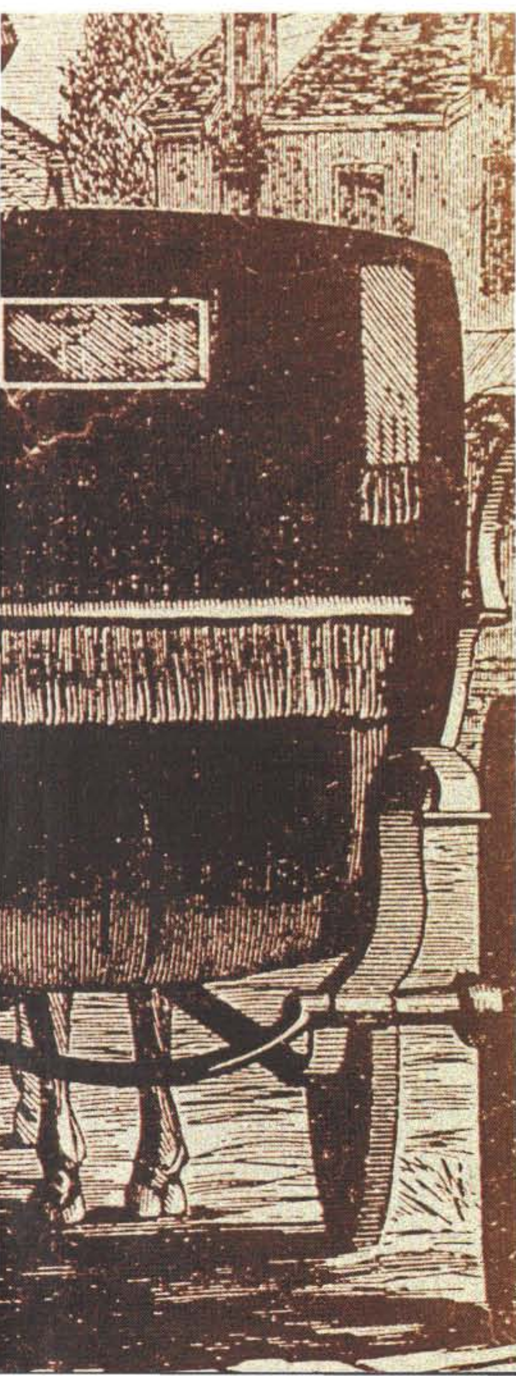
The Philadelphia Yellow Fever Epidemic of 1793

One of the first major epidemics of the disease in the U.S., it devastated America's early capital. It also had lasting repercussions for the city and country

by Kenneth R. Foster, Mary F. Jenkins and Anna Coxe Toogood

Epidemics that quickly wipe out large segments of communities are rare in the U.S. these days. But before the 20th century, such disasters occurred quite frequently, ravaging bewildered populations who usually had little understanding of the causes. Apart from the human tragedies these episodes produced, some had lasting political effects on the nation. A dramatic example arose in 1793, when one of the earliest significant epidemics of yellow fever in the U.S. struck Philadelphia, then the nation's capital.

At the time, Philadelphia was the country's largest and most cosmopolitan city. Yet prominence and prosperity offered little protection. Over the summer and fall, roughly a tenth of its population, some 5,000 people, perished.



CORBIS-BETTMANN

The trouble began soon after French refugees from a bloody slave rebellion in Saint Domingue (now Haiti) landed on the banks of the Delaware River, which bounds the east side of the city. In addition to news of the fighting, the refugees relayed tales of a mysterious pestilential fever decimating several islands in the West Indies. In July the same disease broke out in Philadelphia.

The first to fall were working-class people living along the Delaware. They suffered high fevers and hemorrhages; their eyes and skin turned yellow; and they brought up black vomit. Many died from internal bleeding within days of becoming ill.

These early victims escaped notice by leaders of the medical establishment, perhaps because of their location (both geographic and social) on the margins of the city. By August 19, however, a prominent physician named Benjamin Rush had seen several similar cases and concluded that the patients suffered from "bilious remitting yellow fever." Rush, then in his late 40s, had encountered this malady once before, during his apprenticeship. He immediately warned that an epidemic was under way and took steps to combat it.

Rush stood in good position to influence the city's response to the disaster. He was a renowned professor at the University of Pennsylvania and a founder of the prestigious College of Physicians of Philadelphia. He was famous as a patriot and a signer of the Declaration of Independence, in addition to being a philanthropist and teacher. He was also possessed of enormous energy and a forceful personality.

At Rush's urging, Philadelphia's mayor, Matthew Clarkson, asked the College of Physicians to recommend public health measures against the fever. It issued its report, drafted by Rush, on August 26. Following the prevailing theory that the disease was contagious and spread by putrid vapors, the college told citizens to avoid people stricken with the disease, to breathe through cloths soaked with camphor or vinegar and to burn gunpowder to clear the air. It also proposed establishing a hospital to treat indigent victims, who were too poor to pay for the home care most people preferred. To avoid alarming the public further, the college called for the silencing of church bells, which had been ringing incessantly to announce the many funerals in the city.

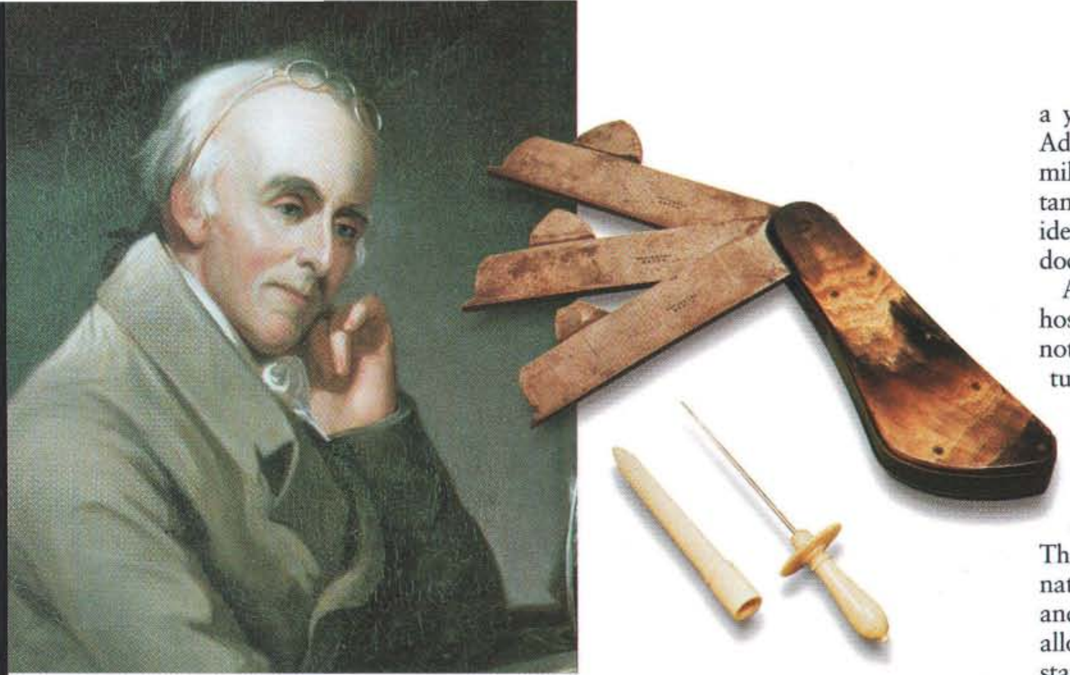
YELLOW FEVER VICTIM is helped to a carriage collecting the dead and dying in an undated woodcut recalling Philadelphia's 1793 epidemic. His aide is a businessman, Stephen Girard. In the background, a man covers his mouth to avoid contracting the disease, then thought to be spread through the air. At the top of this page is *Aedes aegypti*, the mosquito now known to transmit the yellow fever virus.

HARRY ROGERS *Photo Researchers, Inc.*

Yellow fever is now known to be a viral infection transmitted by female mosquitoes of the species *Aedes aegypti*. Evidently, Philadelphia's troubles stemmed from virus-carrying *A. aegypti* stowaways that arrived by ship with the refugees from Saint Domingue. After the infected insects spread the virus to people, uninfected *A. aegypti* picked up the viral agent from the blood of diseased patients and injected it into new victims.

Nevertheless, concern that yellow fever was spread through the air was reasonable in the 18th century, given the state of science at the time. Physicians knew nothing about microorganisms and the transmission of disease, and not until 1900 did Walter Reed and his colleagues finally prove the proposal, originally put forward by Cuban physician Carlos Finlay, that yellow fever was transmitted to humans by mosquitoes. Accordingly, many doctors, including Rush, clung to the medieval theory that diseases were caused by impurities in the air, particularly vapors from decaying vegetable matter.

Proponents of environmental theories of illness had a lot to worry about in Philadelphia. The city lacked an efficient sewage system, and outhouses and industrial waste polluted the water supply. Noxious fumes filled the air from tanneries, distilleries, soap manufacturers and other industries. Animal carcasses lay rotting on the banks of the Delaware and in public streets, particularly in the market along High Street (now called Market Street). Rush himself attributed the disease to vapors from a shipload of coffee that had spoiled en route from the West Indies and lay abandoned and rotting on a wharf. At his urging, Mayor Clarkson republished the city's unenforced laws mandating twice-weekly trash collection and ordered a



BENJAMIN RUSH, a leading physician, fomented controversy during the epidemic by strongly advocating aggressive bloodletting as a "cure" for yellow fever. The instruments shown, from about 1790, are types Rush and his followers might have used. Rush's opponents took a gentler approach, involving rest, cleanliness, wine and Peruvian bark.

CHARLES WILLSON PEALE *Independence National Historical Park (left)*; BRIDGEMAN ART LIBRARY *(right)*

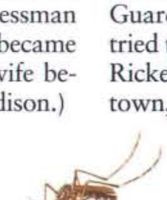
cleanup of the city's streets and markets.

Other doctors, notably William Currie, believed sick refugees from Saint Domingue brought the fever and spread it by physical contact. Those physicians argued for better quarantine regulations. The mayor tried to comply by ordering that arriving passengers and goods be kept isolated for two to three weeks, but during most of the epidemic the city lacked the ability to enforce the order. Even Currie, though, suggested that a combination of very dry weather that summer and putrid vapors had nurtured the disease by destroying the city's "oxygen gas or pure air."

Panic Starts

In correspondence, Rush estimated that 325 people died in August alone. By the end of that month, many people who could afford to leave Philadelphia did so, including some of the city's 80 doctors. (Among the professionals who stayed was John Todd, a lawyer who handled the surge of legal matters arising from the many deaths. Todd caught the fever and died. Within a year, his widow, Dolley, married Congressman James Madison. Madison later became the fourth U.S. president; his wife became the First Lady, Dolley Madison.)

The residents who remained—mostly the poor—followed the advice of the college and applied other tactics to protect themselves. They wore camphor bags or tarred rope around their necks, stuffed garlic into their pockets and shoes, doused themselves with vinegar, shot off guns in their living rooms and lit fires in the streets.



Soon people began abandoning their dying spouses in the streets. Hungry and frightened children roamed the city, their parents dead from the fever. The High Street market stood empty; other commercial activity closed down; and the churches and Quaker meeting houses lost their congregations. Pennsylvania governor Thomas Mifflin departed on September 6, leaving the crisis in Mayor Clarkson's hands. Within a week of Mifflin's exit, nearly all in authority had fled, and the economy had collapsed.

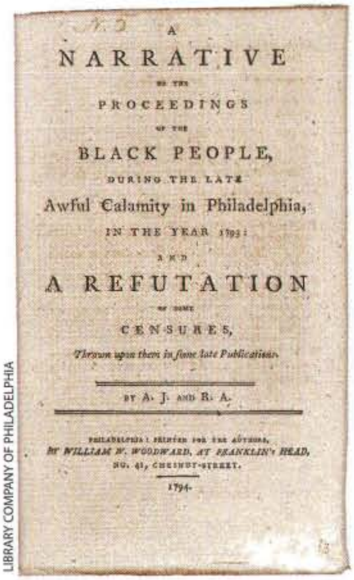
Among Clarkson's top priorities before and after the exodus was providing care for the stricken, although his choices were limited by public fear of contagion. By late August many institutions—among them the Pennsylvania Hospital, the city's almshouse and the Quaker almshouse—had stopped admitting yellow fever patients, to protect current residents. In response the Overseers and Guardians of the Poor (a city agency) tried to appropriate for a hospital John Rickett's circus on the outskirts of town, but angry neighbors threatened to burn the structure down. On August 31 the agency turned to Bush Hill, an unoccupied building tenanted just a year earlier by Vice President John Adams and his wife. Bush Hill stood a mile from the center of town, too distant to elicit resistance from the city residents but also too far for beleaguered doctors to visit regularly.

At first, conditions at the makeshift hospital were abysmal. As one observer noted, "It exhibited as wretched a picture of human misery as ever existed. A profligate, abandoned set of nurses and attendants...rioted on the provisions and comforts prepared for the sick, who... were left almost entirely destitute of every assistance. The dying and dead were indiscriminately mingled together. The ordure and other evacuations of the sick were allowed to remain in the most offensive state imaginable."

On September 10 Mayor Clarkson appealed for volunteers to organize a response to the hospital's and the city's difficulties in caring for the sick. About a dozen men then worked for 46 consecutive days to make the needed arrangements. They borrowed money, bought hospital supplies and other items needed by the poor, rented quarters for an orphanage and hired a staff. And they went door-to-door to check for the dead and dying and to rescue orphaned children. It was an incredible and brave effort in the face of a disease that most doctors believed could be contracted easily by breathing air or by personal contact. Three of those dedicated volunteers eventually died of the fever.

Many African-Americans risked their lives as well. Shortly before the mayor's appeal, Rush had pleaded for nursing assistance from the city's African-American community. Most of its 2,500 members were free (albeit living in poverty) because Pennsylvania was phasing out slavery. Six years earlier two community leaders, Absalom Jones and Richard Allen, had formed the Free African Society, the first self-help group in the U.S. organized by blacks. Rush, an outspoken abolitionist, had close associations with both men and was a strong supporter of the society. From accounts he had read, Rush believed blacks were immune to yellow fever and therefore in a position to help the sick. (In the end, he was proved wrong; more than 300 African-Americans would die in the epidemic, in a proportion close to that of the rest of the population.)

The society agreed to provide workers. Jones, Allen and society member William Gray organized the hiring and



A NARRATIVE OF THE PROCEEDINGS OF THE BLACK PEOPLE, DURING THE LATE Awful Calamity in Philadelphia, IN THE YEAR 1793: AND A REFUTATION OF SOME CENSURES, Thrown upon them in some late Publications.

BY A. J. AND R. A.

PHILADELPHIA: PRINTED FOR THE AUTHORS, BY WILLIAM W. WOODWARD, AT FRANKLIN'S HEAD, NO. 41, CHESNUT-STREET. 1794.

LIBRARY COMPANY OF PHILADELPHIA

RAPHAELLE PEALE *Delaware Art Museum*

ABSALOM JONES, founder with Richard Allen of the Free African Society, worked with Allen and William Gray of the society to recruit African-Americans as nurses for the epidemic. The recruits won praise. But later Mathew Carey, the main historian of the episode, accused some of profiteering and theft. Infuriated, Jones and Allen wrote a rebuttal (*left*) thought to be the first African-American political publication.

dispatching of nurses to homes. Members of the black community also brought people to the hospital and buried the dead. At the society's request, the mayor released black convicts from the Walnut Street prison to work in the "contagious hospital."

On September 18 Rush wrote to his wife (who had left the city with the couple's children), "Parents desert their children as soon as they are infected, and in every room you enter you see no person but a solitary black man or woman near the sick. Many people thrust their parents into the streets as soon as they complain of a headache."

As the nurses recruited by the society began caring for sick people around the city, two exceptional members of the mayor's committee—Stephen Girard and Peter Helm—took charge of Bush Hill. Girard, then a little-known merchant, was a French immigrant who later became famous as the financier of the War of 1812 and the founder of Girard Bank and Girard College for orphan boys. Helm was a cooper of German descent.

Girard accepted responsibility for operating the inside of the building, Helm the exterior and outbuildings. Both showed up every day until the epidemic finally waned later in the year. Girard established order and cleanliness and made sure the patients were given careful attention. Helm restored the water supply to the grounds, built coffins and more hospital space and supervised the reception of patients. This reception included placing the patients in holes in the ground to protect them from vapors while they waited for hospital beds.

Girard was familiar with yellow fever through trading in the West Indies. He persuaded the mayor's committee to hire the physician Jean Devèze, one of the refugees from Saint Domingue. Devèze had treated the disease while in the French army and would become the world's leading authority on the condition. He advocated the so-called French method of therapy: bed rest, cleanliness, wine and treatment with Peruvian bark (a source of quinine, which is now used to treat malaria, not yellow fever). At Bush Hill, Girard generously supplied the wine from his own cellar.

COLLEGE OF PHYSICIANS OF PHILADELPHIA

LATE SIGNS of yellow fever, such as a profusely bloody nose and black vomit (*seen on pillow*), were depicted in 1820 as part of a series illustrating the course of the disorder. The series, from a French publication, includes some of the earliest color images of the disease's stages.

Controversy over Treatments

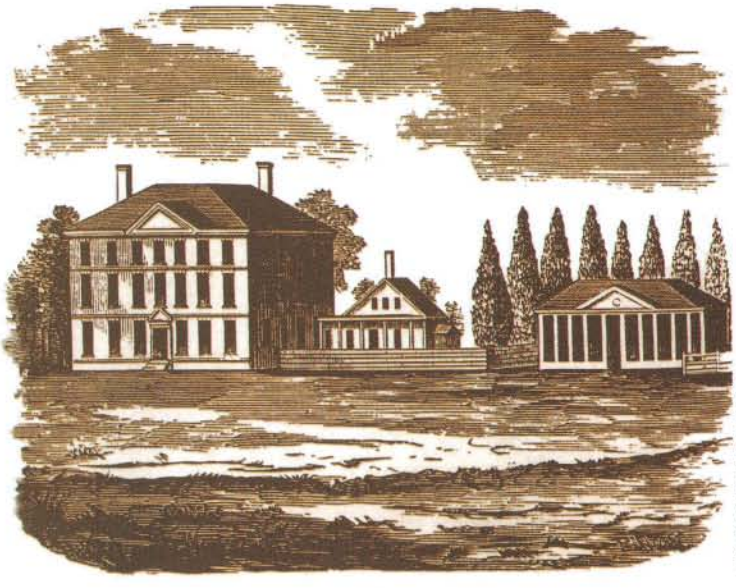
American physicians objected to Devèze's appointment, however, and four resigned from the hospital. In part, they may have been angry at losing medical authority over Bush Hill. Also, they were trained in the Scottish and English tradition, which was quite different from that of the French. Indeed, Devèze stood at the opposite pole from Rush and his supporters in what became a raging controversy over the best therapy.

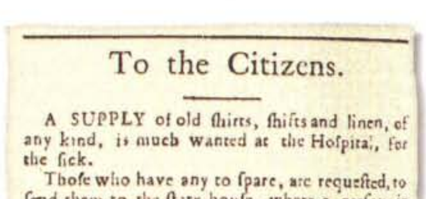
In accordance with traditional teachings, Rush and many of his contemporaries believed the body contained four humors (blood, phlegm, black bile and yellow bile). They therefore aimed to treat most illnesses by restoring balance to the body—through giving laxatives and emetics (to cause vomiting), drawing blood and inducing sweating.

Early on, Rush had become convinced that a combination of these treatments could effect a cure—specifically, he favored bleeding combined with delivery of a mercury-based mixture of calomel and jalap. "I preferred frequent and small, to large bleedings in the beginning of September," he wrote in a 1794 account of the epidemic, "but toward the height and close of the epidemic, I saw no inconvenience from the loss of a pint, and even 20 ounces of blood at a time. I drew from many persons 70 and 80 ounces in five days, and from a few, a much larger quantity."

Over time, Rush became increasingly dogmatic. He claimed he "did not lose a single patient" who had been bled seven times or more. He also mounted a strenuous campaign, in private correspondence and in letters to the editors, urging doctors not to spare the lancet. His letters fiercely criticized physicians who seemed to pay lip service to the benefits of venisection but were more restrained than him in its practice. He aimed particularly strong invective at Devèze and the few other doctors who thought that radical bloodletting was dangerous and said so openly.

As Rush's treatment became progressively more radical, his medical colleagues began to drop their loyalty. But prominent Philadelphians still often


LIBRARY COMPANY OF PHILADELPHIA


To the Citizens.

A SUPPLY of old ſhirts, ſhifts and linen, of any kind, is much wanted at the Hoſpital, for the ſick.

Thoſe who have any to ſpare, are requeſted, to ſend them to the ſtate-houſe, where a perſon is appointed to receive them.

MATTHEW CLARKSON,
Mayor

LIBRARY COMPANY OF PHILADELPHIA

BUSH HILL (*top*), where Vice President John Adams and his wife had lived for a time, became a hospital for indigent yellow fever victims. To aid the hospital, the mayor published a call requesting donated supplies (*bottom*).

turned to Rush for care. John Redman, Rush's mentor and still at 71 years the reigning dean of medicine in Philadelphia, submitted to Rush's therapy and lived. Caspar Wistar, a respected professor of anatomy, recovered, too. In Rush's immediate circle, three of his apprentices died, as did his own sister. Rush and two other apprentices contracted yellow fever but recovered after taking his treatment.

Historians know that neither Rush's method nor Devèze's could have cured patients. With or without intervention, yellow fever runs a variable course, killing some but sparing others. As is true of most viral diseases, it remains essentially untreatable today, and therapy mainly consists of keeping patients comfortable and preventing dehydration. Devèze's gentle approach certainly did less harm than Rush's and may have helped patients remain strong enough to combat the virus. But it would be unfair to judge Rush for his decisions. His protocol was simply an aggressive version of a conventional therapy based on the medical theories of the day.

The epidemic died out on its own in November, when frosts killed the mosquitoes. The city kept no accurate health records, so the precise number of deaths cannot be known. By counting fresh graves in the city's cemeteries and examining church records and other sources, Mathew Carey, who became the official historian of the epidemic, reported the toll as 4,041 out of a total population of 45,000. The actual number was surely higher, however. In an account published within months after the epidemic, even Carey suggested that the figure exceeded 5,000.

Of the physicians who remained in town, 10 paid for their dedication with their lives. Overall, as many as 20,000 residents fled, a number of them ending up quarantined or jeered by frightened mobs along the roads out of town.

Several yellow fever epidemics were to follow in Philadelphia, in 1797, 1798, 1799, 1802, 1803 and 1805. Epidemics also hit other parts of the U.S. at various times, especially in the South, where *A. aegypti* is common. The last major outbreak occurred in New Orleans in 1905.

For Rush, the 1793 event represented both the apogee and nadir of his career. Initially, his dedication to patients made him a popular hero; he reportedly saw 120 patients a day. But he was too sure of himself, intolerant of alternative therapies and unable to bear criticism. By imposing his will and politicizing the medical community, he became a formidable obstacle to more enlightened treatment by other physicians. "The different opinions of treatment excite great inquietude," wrote Henry Knox, secretary of war, to President George Washington, on September 15, 1793. "But Rush bears down all before him."

Ultimately, Rush's divisiveness and radical treatment tarnished his reputation. Enraged by perceived slights from other doctors, he resigned from the College of Physicians shortly after the epidemic. Unbowed, however, he dedicated himself to treating victims of the yellow fever epidemics that struck Philadelphia subsequently, practicing even more radical bloodletting than before. In 1797 his loyal former student, Philip Physick, survived 22 bleedings and reportedly lost 176 ounces of blood.

For Philadelphia, the 1793 epidemic led in the following year to creation of the city's first health department, among the earliest in the nation. And partly as a result of the epidemic, in 1801 Philadelphia completed the first municipal water system in a major American city. Health authorities assumed eliminating filth would help prevent yellow fever and other diseases; building a water system was simply part of that general cleanup.

The efforts of the African-American community improved relations between blacks and whites in Philadelphia when racial tension in much of the country was mounting. During the epidemic, the steadfastness of the blacks as nurses and handlers of the dead won widespread recognition. In 1794 gratitude led city leaders and an initially resistant white clergy to cooperate with the African-American community as it founded the first black-owned and -operated churches in the nation. Those churches, African Episcopal Church of St. Thomas and Mother Bethel A.M.E. Church, still have active congregations today.

Broad Repercussions

All was not smooth, though. Historian Carey accused some blacks of exploiting the situation, charging gouging prices for their services and pillaging the homes of patients they had cared for. In what some experts think is the nation's earliest African-American political publication, Jones and Allen published an angry but restrained rebuttal, vividly describing the horrible conditions under which the nurses worked, unassisted and sometimes refusing payment for their services.

The repercussions of the 1793 epidemic extended beyond Philadelphia. During that episode, Ann Parish, a Quaker woman of independent means, established an institution that enabled women widowed by the disease to keep their families intact. (While the women spun wool at the House of Industry, others watched and taught their children.)

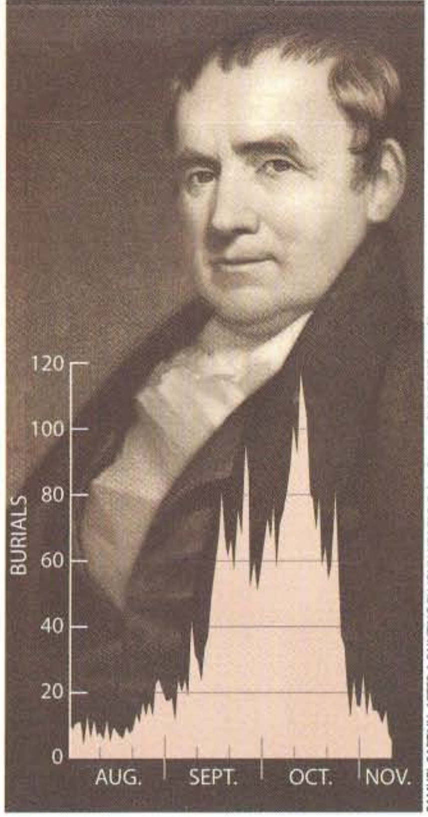
Her program became a model of innovative philanthropy for other cities that later suffered yellow fever epidemics. Meanwhile, as a result of the controversy stirred up by Rush, French methods of treatment became better known among doctors elsewhere, many of whom now debated the merit of extreme bloodletting.

Along the seaboard, cities developed measures for public health, including establishing quarantine stations for foreign ships. Moreover, a sanitation movement began that carried into the 19th century. For example, various cities built water and sanitation systems and cleaned the streets more often.

The epidemic—occurring as it did in the nation's capital at a crucial time in the country's history—had important political ramifications as well. Before yellow fever arrived, public sentiment in the capital leaned strongly toward supporting the revolutionary government of France in the war it had declared against Great Britain, the Netherlands and Spain. An envoy sent by France to stir up popular support for the war, Citizen Edmond Charles Genêt, found enthusiastic crowds in Philadelphia but received a cool reception from President Washington.

John Adams, who would become second president of the U.S., later remembered that "ten thousand people in the streets of Philadelphia day after day, threatened to drag Washington out of his house, and effect a revolution in the government, or compel it to declare war in favor of the French Revolution, and against England.... Nothing but the yellow fever... could have saved the United States from a total revolution of government."

The medical catastrophe defused the political excitement. By the time the epidemic ended, news of the excesses of the French government had reached America and public opinion had tilted against France, toward neutrality.



SAMUEL SARTAIN, AFTER A PAINTING BY JOHN NEAGLE *Culver Pictures*; SLIM FILMS *(graph)*

HISTORIAN MATHEW CAREY calculated lives lost to yellow fever in various ways, such as by counting new graves in cemeteries (*graph*). Such methods suggested a total of 4,041. Later estimates, however, put the toll closer to 5,000.

After reappearing many times, yellow fever gradually subsided in the U.S. late in the 19th century. It nonetheless continues to threaten certain tropical areas, particularly in South America and sub-Saharan Africa.

The decline in the U.S. began long before an effective vaccine was introduced in the 1930s. (The vaccine is available for soldiers and travelers; unfortunately, it costs too much for populations living in the areas most prone to epidemics today.) Quarantine measures and municipal water supplies that reduced breeding areas for mosquitos probably played a role in ending the U.S. epidemics, but other factors, such as improved sanitation, must have been at work as well. The drop-off is part of a larger pattern in which death rates from many infectious diseases fell in the course of the 1800s.

In tropical regions, the virus finds a natural reservoir in monkeys, and there is no hope of eliminating it by destroying its natural hosts. Scientists may eventually be able to control the disease by displacing the native mosquito population with genetically altered forms that cannot spread the disease, but that approach is far from realization. Meanwhile a yellow fever epidemic remains a worrisome possibility in the U.S., particularly in the Deep South, where large populations of *A. aegypti* are found.

In spite of modern knowledge and the existence of a vaccine, such an epidemic could grow quickly. Once it began, public health workers would need time to immunize populations at risk and to establish effective mosquito-control and quarantine measures. The U.S. Institute of Medicine has estimated that an outbreak of yellow fever in a major city, such as New Orleans, would potentially result in as many as 100,000 cases and 10,000 deaths.

The institute has additionally predicted that yet unknown viruses could also become major health threats. (HIV, the virus that causes AIDS, is an example of a once obscure virus that triggered a major epidemic.) It has therefore urged the government to adopt better methods for identifying new outbreaks of infectious diseases, to stop them early. After their experience with the 1793 yellow fever epidemic in Philadelphia, Benjamin Rush and his medical colleagues would have firmly agreed. SA

The Authors

KENNETH R. FOSTER, MARY F. JENKINS and ANNA COXE TOOGOOD all share an intense interest in Philadelphia's history. Foster is associate professor of bioengineering at the University of Pennsylvania. Jenkins is supervisory park ranger at Independence National Historical Park in Philadelphia. Toogood has been a historian in the Park Service for more than 30 years. Foster and Jenkins are married to each other.

Further Reading

A NARRATIVE OF THE PROCEEDINGS OF THE BLACK PEOPLE DURING THE LATE AWFUL CALAMITY IN PHILADELPHIA IN THE YEAR 1793, AND A REFUTATION OF SOME CENSURES THROWN UPON THEM IN SOME LATE PUBLICATIONS. Absalom Jones. Historic Publications, Philadelphia, 1969. Reprinted by Independence National Historical Park, Philadelphia.

FORGING FREEDOM: THE FORMATION OF PHILADELPHIA'S BLACK COMMUNITY 1720–1840. Gary B. Nash. Harvard University Press, 1988.

BRING OUT YOUR DEAD: THE GREAT PLAGUE OF YELLOW FEVER IN PHILADELPHIA IN 1793. J. H. Powell. Reprint edition. University of Pennsylvania Press, 1993.

A MELANCHOLY SCENE OF DEVASTATION: THE PUBLIC RESPONSE TO THE 1793 PHILADELPHIA YELLOW FEVER EPIDEMIC. Edited by J. Worth Estes and Billy G. Smith. Science History Publications/USA, Philadelphia, 1997.

THE AMATEUR SCIENTIST

by Shawn Carlson

Building a Consciousness of Streams

Okay, I'll confess: my scientific specialty is nuclear physics, but I would much rather trace the course of a quiet stream than find my way along a linear accelerator any day. And ecology, especially stream ecology, affords more opportunities for meaningful amateur research. Indeed, the techniques used are so straightforward and the results so immediate that you can easily craft wonderful scientific experiences for your whole family with just a little preparation. And you can forge ties with others, too—with the rise of the environmental movement, thousands of part-time naturalists have banded together into nearly 800 independent organizations that work to assess the health of our nation's watersheds. These associations can help you and your family jump into this rich and rewarding area of research.

Some of these groups have just a few members and operate on a shoestring. Others have hundreds of participants and annual budgets of tens of thousands of dollars. Some limit their activities to physical and biological surveys: they record the temperature, rate of flow and clarity of the water, or they take a regular census of the seasonal inhabitants of a stream. Other groups have the equipment to measure dissolved oxygen and pH levels, check for fecal coliform bacteria and test for various pollutants. All these volunteers make important contributions to bettering the environment, albeit one stream at a time.

SAMPLING FAUNA provides a simple test for pollution.

These organizations report their findings to local, state and federal agencies that have an interest in water quality. The Environmental Protection Agency encourages this burgeoning torrent of amateur talent with a number of free publications. The EPA also helps to produce an outstanding semiannual newsletter called *The Volunteer Monitor.* This periodical knits independent environmental groups into a nationwide community by providing detailed instructions, some that will enlighten even the most experienced field researcher. It's easy to get your feet wet: just check out the EPA's listings on the World Wide Web to find a watershed volunteer organization near you.

As with any diverse ecosystem, streams afford many opportunities for amateur research. Personally, I love to investigate the little critters that cling to rocks and branches or bury themselves in the sediment. Studying the diversity of these easily visible "macroinvertebrates" provides an excellent way to judge the health of the stream, because pollution affects certain types more than others. For example, finding stone-fly nymphs (relatively sensitive creatures) inhabiting a brook in one of the Middle Atlantic states would signal that the level of pollution is probably not too severe. But the rules change from place to place. So you will need to learn from people working in your area which of the local organisms are most sensitive to pollution.

To examine such invertebrates, you may find that some of the methods used for standing bodies of water are helpful


RICK JONES

[see "The Pleasures of Exploring Ponds," The Amateur Scientist, September 1996]. But the rush of the relatively shallow water in streams makes the technique described here particularly well suited to collecting biological specimens.

To start, you'll need a good net. Affix a one-meter square of fine-mesh netting between a pair of strong supports. (Stout wooden dowels would work well.) Environmental groups in your area may use a standard-size mesh for such studies—500 microns is common—in which case you should do the same. Otherwise you can just use plastic window screening. Then go out and select three representative sites from a section of your stream that is no more than 100 meters long. Your aim is to sample a one-meter-square patch in each spot. To do so, you and a partner should put on rubber hip boots and wade in, approaching from downstream so that anything kicked up by your steps will not affect your investigation.

When you reach your first site, open the net, dig the ends of the dowels into the bottom so that the net tilts downstream at a 45-degree angle. Make sure the base of the net lies flush against the streambed. Have your helper step one meter in front of the net and pick up all the large rocks from the sampling area, gently rubbing them underwater to dislodge any organisms living on them. The flow of the stream will carry anything rubbed off into the net. (Alternatively, rubbing the rocks over a bucket may be a more direct way for you to collect these organisms.) Your partner should then stir up the bottom in the designated patch by kicking things around for a couple of minutes so that the organisms living between the smaller rocks are swept into the net. Finally, have your assistant grasp the bottom of the net and scoop it forward as both of you remove it from the stream and return to shore.

Use a magnifying glass to find and collect creatures from twigs and whatever other debris you've captured before you discard them back into the stream. Carefully dislodge the larger creatures from the mesh with tweezers; flush off the rest into a bucket with clean stream water. Let the contents of the pail settle and scoop out most of the excess water with a cup. Strain this water through a nylon stocking to make sure that nothing escapes. Then repeat this procedure at your other two survey sites, combining the specimens in your bucket.

Pour the water remaining in the bucket through your stocking strainers. Next, turn them inside out and use a little rubbing alcohol to wash the contents into an empty glass canning jar. Use a sugar scoop and tweezers to place the rest of the pail's living contents into the jar and fill it with alcohol to kill and preserve all these invertebrates before screwing on the top. Finally, place a large label on the jar and pencil onto it (alcohol causes ink to run) the location, date, time and the names of everyone on your crew.

Later you can go through the collection and record the types and quantities of organisms present. Environmental groups working in your area can help you learn how to identify the sundry invertebrates. Or, at the very least, they should be able to point you in the right direction for gaining these skills. With this knowledge, you can revisit these same sites periodically to document environmental changes from season to season and, eventually, from year to year.

Although you'll probably want to gain some practice with streams close to your home, don't overlook the bigger picture. You can see how your stream fits into your regional river system by getting a 7.5-minute (scale 1:24,000) topographic map by the U.S. Geological Survey. You can purchase such maps at hiking supply stores or directly from the USGS. By checking with other groups and plotting the existing survey sites on your map, you'll quickly discover those places in need of study. Then grab your net, your bucket, your boots and a few family members and go to it. SA

For more information about this and other projects for amateurs, visit the Forum section of the Society for Amateur Scientists at http://web2.thesphere.com/SAS/WebX.cgi on the World Wide Web. You may also write the Society for Amateur Scientists, 4735 Clairemont Square, Suite 179, San Diego, CA 92117, or call 619-239-8807.

RESOURCES

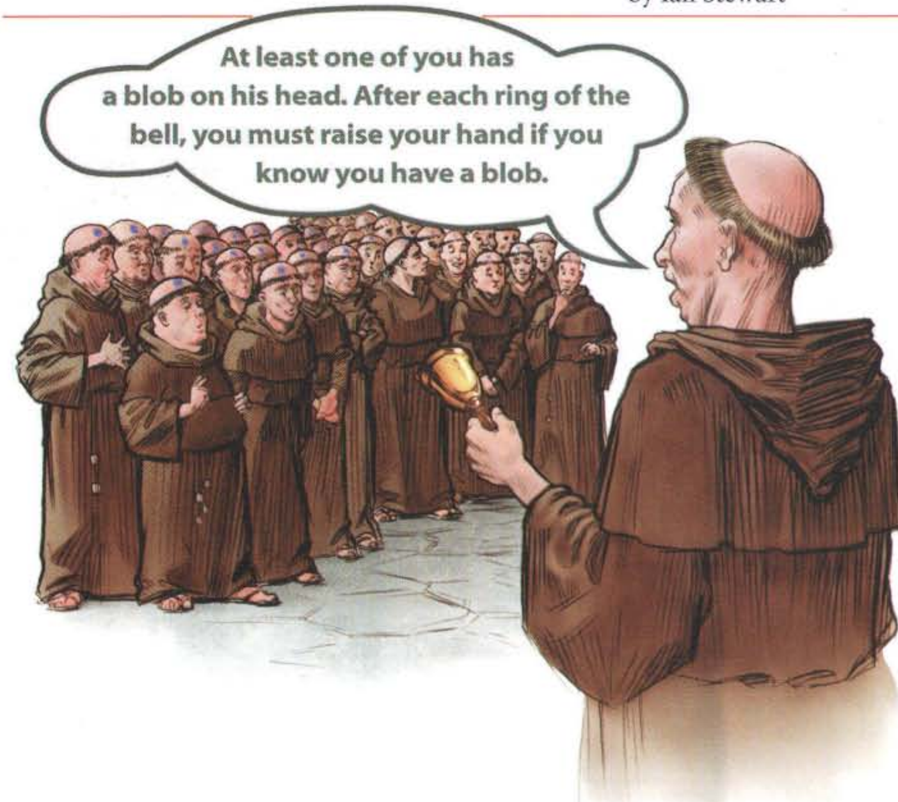
The pamphlet "EPA's Volunteer Monitoring Program" is available from USEPA, Volunteer Monitoring, 401 M Street SW (4503F), Washington, D.C. 20460. For information, check out http://www.epa.gov/owow/monitoring/vol.html on the World Wide Web.

To obtain the national newsletter, send your name and address to *The Volunteer Monitor*, 1318 Masonic Avenue, San Francisco, CA 94117, or call 415-255-8049.

To learn how to purchase a map from the USGS, call 800-USA-MAPS.

Seniors looking for volunteer opportunities should contact the Environmental Alliance for Senior Involvement (EASI). Write to EASI, 8733 Old Dumfries Road, Catlett, VA 20119, or consult http://www.easi.org on the World Wide Web.

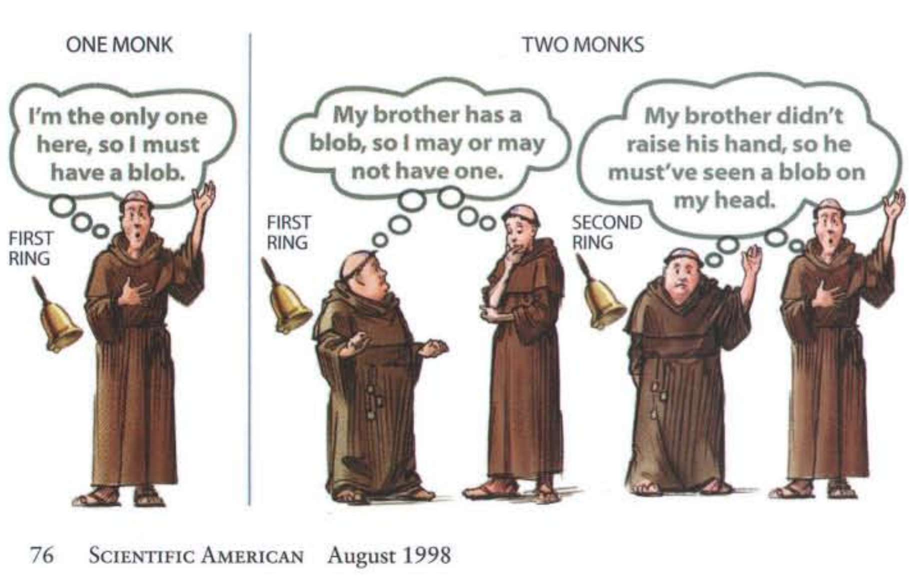
by Ian Stewart



Monks, Blobs and Common Knowledge

The extremely polite monks of the Perplexian order like to play logical tricks on one another. One night, while Brothers Archibald and Benedict are asleep, Brother Jonah sneaks into their room and paints a blue blob on each of their shaved heads. When they wake up, each notices the blob on the head of the other but, being polite, says nothing. Each vaguely wonders if he, too, has a blob but is too polite to ask. Then Brother Zeno, who has never quite learned the art of tact, enters the cell and begins giggling. On being questioned, he says only, "At least one of you has a blue blob on his head."

Of course, both monks already know that. But then Archibald starts thinking: "I know Benedict has a blob, but he doesn't know that. Do I have a blob? Well, suppose I don't have a blob. Then Benedict will be able to see I don't have a blob and will immediately deduce from Zeno's remark that he must have a blob. But he hasn't shown any sign of embarrassment—oops, that means I must have a blob!" At which point, he blushes bright red. Benedict also blushes at about the same instant, for much the same reason. Without Zeno's remark, neither train of thought could have been set in motion, yet Zeno tells them nothing—apparently—that they do not know already.



ILLUSTRATIONS BY MATT COLLINS

When you start to analyze what's going on, however, it becomes clear that Zeno's announcement—"At least one of you has a blue blob on his head"—does indeed convey new information. What do the monks actually know? Well, Archibald knows that Benedict has a blob, and Benedict knows that Archibald has a blob. Zeno's statement does more than just inform Archibald that someone has a blob—it also tells Archibald that Benedict now knows that someone has a blob.

There are many puzzles of this kind, including versions involving children with dirty faces and partygoers wearing silly hats. These logical conundrums are called common knowledge puzzles because they all rely on a statement known to everyone in the group. It is not the content of the statement that matters: it is the fact that everyone knows that everyone else knows it. Once that fact has become common knowledge, it becomes possible to reason about other people's responses to it.

This effect gets even more puzzling when we try it with three monks. Now Brothers Archibald, Benedict and Cyril are asleep in their room, and Jonah paints a blue blob on each of their heads. Again, when they wake up, each of them notices the blobs on the others but says nothing. Then Zeno drops his bombshell: "At least one of you has a blue blob on his head."

Well, that starts Archibald thinking: "Suppose I don't have a blob. Then Benedict sees a blob on Cyril but nothing on me, and he can ask himself whether he has a blob. And he can reason like

this: 'If I, Benedict, do not have a blob, then Cyril sees that neither Archibald nor Benedict has a blob and can deduce immediately that he himself has a blob. Because Cyril has had plenty of time to work this out but remains unembarrassed, then I, Benedict, must have a blob.' Now, because Benedict has also had plenty of time to work this out but remains unembarrassed, then it follows that I, Archibald, must have a blob." At this point, Archibald turns bright red—as do Benedict and Cyril, who have followed a similar line of reasoning.

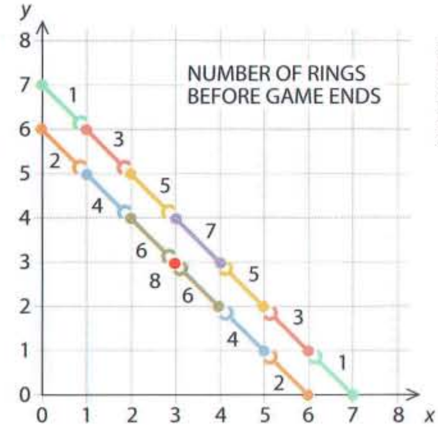
The same logic works with four, five or more monks, although their deductions become progressively more convoluted. Suppose there are 100 monks, each bearing a blob, each unaware of that fact and each an amazingly rapid logician. To synchronize their thoughts, suppose that the monastery's abbot has a bell. "Every 10 seconds," the abbot says, "I will ring this bell. Immediately after each ring, all of you who can deduce that you have a blob must put your hands up. At least one of you has a blob." Nothing happens for 99 rings, and then all 100 monks simultaneously raise their hands after the 100th ring.

The logic goes like this: If there is only one monk in the group, he immediately deduces that the blob must be on his head, so he raises his hand after the first ring of the bell. If there are two monks, each assumes he doesn't have a blob, so neither raises his hand after the first ring. But then each monk deduces from the other's reaction that his assumption is wrong. "If there was no blob on my head," each thinks, "my brother would have raised his hand. Because he didn't, I must have a blob." So they both raise their hands after the second ring. The pattern is the same for any number of monks—it's an instance of mathematical induction, which says that if some property of numbers holds when $n = 1$ and if its validity for n implies its validity for $n + 1$, then it must be valid for all n. If there are 100 monks in the group, each assumes that he has no blob and expects the 99 other monks to raise their hands after the 99th ring. When that doesn't happen, each monk realizes his assumption was wrong, and after the next ring everyone raises his hand.

Another intriguing common knowledge puzzle was invented by John H. Conway of Princeton University and Michael S. Paterson of the University of Warwick in England. Imagine a Mad Mathematicians' tea party. Each partygoer wears a hat with a number written on it. That number must be greater than or equal to 0, but it need not be an integer; moreover, at least one of the players' numbers must be nonzero. Arrange the hats so no player can see his own number but each can see everyone else's.

Now for the common knowledge. Pinned to the wall is a list of numbers, one of which is the correct total of all the numbers on the players' hats. Assume that the number of possibilities on the list is less than or equal to the number of players. Every 10 seconds a bell rings, and anyone who knows his number—or equivalently knows the correct total, because he can see everyone else's numbers—must announce that fact. Conway and Paterson proved that with perfect logic, eventually some player will make such an announcement.

Consider just two players, with the numbers x and y on their hats, and suppose that the list pinned to the wall has the numbers 6 and 7. Both players know that either $x + y = 6$ or $x + y = 7$. Now for some geometry. The pairs (x,y) that satisfy these two conditions are the coordinates of two line segments in the positive quadrant of the plane [*see illustration above*]. If x is greater than 6, the y player will end the game after the first ring of the bell, because he can see immediately that a total of 6 is impossible. Similarly, the x player will end the game



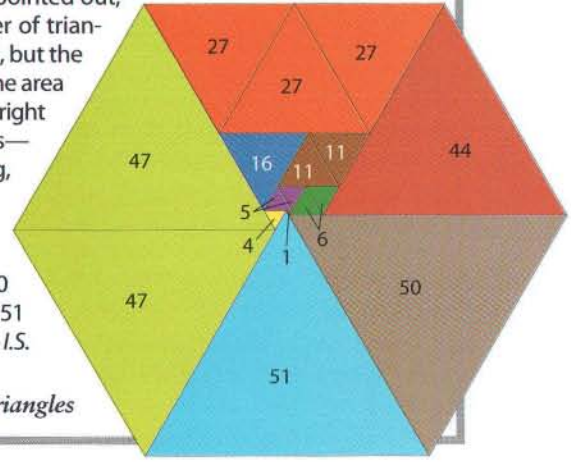
LINE SEGMENTS show how long the two-player game lasts.

if y is greater than 6. If neither player responds after the first ring, these possibilities are eliminated. The game then terminates on the second ring if either x or y is less than 1. Why? One of the players can see the hat with a number less than 1 and knows that his own number must be 6 or less; therefore, the total of 7 is ruled out.

With each ring of the bell, the pairs (x,y) that are eliminated form successive diagonal segments of the two original line segments, thereby quickly exhausting all the possibilities. The game must end by the eighth ring if x and y both equal 3. Every other possibility requires seven or fewer rings. A similar argument can be applied to games with three or more players, although the proof is more mathematically sophisticated. SA

FEEDBACK

Feedback in the July 1997 column discussed a problem suggested by Robert T. Wainwright of New Rochelle, N.Y., asking for the largest possible area that can be covered by equilateral triangles whose sides are integers that have no common divisor. As several readers pointed out, there is no upper limit if the number of triangles can be made as large as required, but the point of the problem is to maximize the area for a given number of triangles. Wainwright reached an area of 496 for 11 triangles—where, for convenience in counting, the unit of area is that of an equilateral triangle with a side of 1. Robert Reid of London found tilings with 12 through 17 triangles. The areas are: 860 (12 triangles); 1,559 (13); 2,831 (14); 4,751 (15); 8,559 (16); and 14,279 (17). —*I.S.*



LARGEST TILING for 17 triangles

REVIEWS AND COMMENTARIES

AMERICA'S INDUSTRIAL COMEBACK

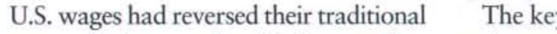
Review by Lewis M. Branscomb

The Productive Edge: How U.S. Industries Are Pointing the Way to a New Era of Economic Growth
BY RICHARD K. LESTER
W. W. Norton and Company, New York, 1998 ($29.95)

One decade ago the Massachusetts Institute of Technology assembled a Commission on Industrial Productivity to examine the reasons American dominance of high-technology industry was experiencing a serious competitive decline. The commission's 1989 book, *Made in America: Regaining the Productive Edge,* documented the problems in a number of industries—problems that were evidenced by poor product quality, noncompetitive production costs and excessively long product cycles, as compared with best practice in Asia. The most painful part of this realization was that the decline affected not only steel, textiles and automobiles but also the high-tech microelectronics, computer and communications industries. This book was a call to arms. What were Americans doing wrong? How could we fix it?

Now the leader of the team that wrote *Made in America* has come out with a sequel, borrowing the title from the subtitle of the earlier study. But the question now being asked is not what did we do wrong, but what are we doing right? Lester examines the easy answers and finds some truth in all of them: our most threatening competitor, Japan, was struggling with the rupture of the "bubble" economy and no longer had almost unlimited access to capital. After the profligate public borrowing of the Reagan administration, politicians of both parties committed themselves to bringing the federal budget into balance. Exchange rates had altered dramatically throughout the 1980s, with the U.S. dollar losing half its value against the yen. And most important—and clearly a matter of serious concern for the future—U.S. wages had reversed their traditional course. Union membership in manufacturing was declining; firms were reducing costs by laying off middle managers who could find only lower-paying jobs.

But these answers, which imply that the favorable conditions industry now enjoys may not last forever and may not be acceptable to the population at large, failed to explain the steps managers had taken to change corporate culture. Lester looks into corporate experience with



SOURCE: Council on Competitiveness

MARKET SHARE of global semiconductor industry for Japan and the U.S., 1982–2000

the "marginal manifestos" offered by an explosion of business management books. He evaluates two approaches that did attempt to address the total performance of the enterprise: Total Quality Management (TQM) and Corporate Reengineering. About two thirds of some 500 firms surveyed by Arthur D. Little were disappointed in these approaches. Lester notes that many firms did not appreciate how profound change had to be and how long it would take for the benefits to be evident.

The truth can be best understood in the strengths and weaknesses of American culture: our tendency to complacency and hubris (my word, not Lester's) and our ability to accept challenges when they finally get our attention and to come up with innovative solutions. American managers seemed slow to react when high-tech trade started to go south around 1970. But once the erosion of U.S. market shares got their attention, many managers brought a level of flexibility and innovation to both organization and strategy that allowed them to respond much more quickly to competitive threats.

The key elements of success are basic and easy to understand. They include loyalty, focus, innovation and a willingness to take risks. But they must be deeply and comprehensively embedded in corporate culture. This explains, at least in part, why it took so long for the turnaround to produce visible results at the macroeconomic level (although some firms, of course, were able to transform their behavior earlier). Once American industry had broken away from the mass-production model and given up hope that exposing everyone to a TQM class would do the trick, the rate of recovery, assisted by financial troubles in Japan, has astonished almost everyone.

The groundwork for the recovery was in large measure already coming into place before the 1990s. Many firms had placed their priorities where they saw the most entrepreneurial opportunity, whereas Asian strategies aimed at commoditizing products to leverage their superior handling of process technology, factory management and cost. While Japan and Korea were placing their bets on commodity memory chips and were failing to master the software business, American entrepreneurship was alive and well. Engineers were leaving semiconductor companies and starting application-driven businesses that could add more value to the electronic hardware.

The U.S. software industry not only grew in size but began to take from hardware the role of defining and delivering function for the user (witness the extraordinary position now occupied

by Microsoft). Businesses paid new attention to total customer satisfaction by delivering a complete system of hardware, software and service. Like IBM, many discovered that offering service may bring more profit to the firm and more value to the customer than simply selling hardware whose price continues to fall. But until companies began to pay attention to their product quality and cost and recognized that a high-tech product built in a low-tech factory could not compete, the more value-added strategy of American firms could not show through in the economic statistics.

Another important factor in the rebirth of manufacturing competitiveness is the power of information technology, which has allowed everyone in the firm to seize the initiative from management and permitted flatter, more flexible organizations. But the Internet, like TQM and Reengineering, proved to be no silver bullet. Its contribution to management and to new business innovation is real and highly significant. But the long time required to realize the benefits of information technology simply points up the rightness of Lester's focus on culture change. A Harvard University study of supply-chain integration in the automobile industry revealed that the automaker and its primary suppliers had great difficulty using well-known computer design and networking technology to integrate their product development until they were able to create a project-driven, collaborative culture.

In short, there is no "one size fits all" solution for U.S. manufacturing. Lester notes in six industry case studies that each industry, and often individual firms, found unique solutions to the risks produced by more open global markets and ever more sophisticated and determined competition. Furthermore, he is not fully satisfied that industry's problems are all solved or that Americans would be satisfied with the solutions. He leaves us with two problems.

Lester dares to ask the hard question: Is the turnaround an illusion? He observes that manufacturing productivity stopped growing at about 2 percent a year around 1973 and has been stuck at less than 0.5 percent ever since. Is this a reflection of the difficulty economists face in measuring productivity? the result of the admixture of services and manufacturing as firms become more responsive to customer needs? or the result of a more competitive economy in which margins continue to be pressed by the most productive performers, and the workers laid off by one firm struggle to find productive roles in another? The optimists among us might feel that as long as the stock market soars, high-tech market shares are growing again and the federal budget is in balance, we can tolerate a lagging productivity rate.

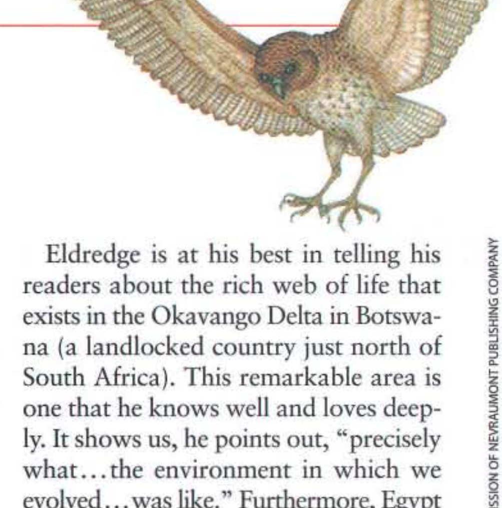
But Lester doesn't leave us much room for a rebirth of complacency. He finds that much of our business success has been bought at the expense of an inexorable growth in income disparity in the American workforce and a deep sense of personal vulnerability to the uncertainties of a market-driven, global economy. He worries that the massive research investments made by government in the past, which provided the intellectual resources for American innovation, may be curtailed. He recognizes that the firms exhibiting best practice are not typical. He is especially concerned about low savings rates and rates of investment. But underlying all the threats and uncertainties about the future is a pervasive empowerment of individuals, together with a great increase in the demand for personal responsibility. Lester asks us to think about how this call for a "new economic citizenship" can enable a continuation of economic success and at the same time reduce the income disparities and sense of vulnerability that deprive many Americans of the full fruits of that success.

LEWIS M. BRANSCOMB, former vice president and chief scientist of IBM, is the Aetna Professor, emeritus, in Public Policy and Corporate Management at the John F. Kennedy School of Government at Harvard University.

LESSONS FROM PARADISE

Review by Peter H. Raven

Life in the Balance: Humanity and the Biodiversity Crisis
BY NILES ELDREDGE
ILLUSTRATIONS BY PATRICIA WYNNE
A Peter N. Nevraumont Book/Princeton University Press, Princeton, N.J., 1998 ($24.95)



PATRICIA WYNNE, USED BY PERMISSION OF NEVRAUMONT PUBLISHING COMPANY

No institution in the U.S. has been more influential in inspiring people of all ages with the wonders of life on Earth, past and present, than the American Museum of Natural History (AMNH) in New York City. Most appropriately, the museum has now turned its attention to biodiversity and opened a splendid new exhibit on biodiversity and the ways in which we are affecting it. Simultaneously, noted AMNH scientist Niles Eldredge has presented a book to enhance the experience and to make more information available. This is a book for the general reader who wants to understand what the mysterious notion of "biodiversity" comprises and to appreciate the damage we are doing to our life-support systems on a daily basis.

Eldredge is at his best in telling his readers about the rich web of life that exists in the Okavango Delta in Botswana (a landlocked country just north of South Africa). This remarkable area is one that he knows well and loves deeply. It shows us, he points out, "precisely what...the environment in which we evolved...was like." Furthermore, Egypt was like this at the time of the pharaohs, and our civilization grew and became more complex and varied under such conditions. Mummified remains of sacred ibises and inscriptions on the walls of the pyramids attest to the accuracy

of this view. Several thousand years ago droves of wildlife and all the interwoven ecological systems of which they were part existed throughout the East African Rift Valley and, though sometimes in diminished expression, far beyond.

Now they exist only in the Okavango. There our ancestors lived in balance with and exploited the productivity of the ecosystems that still occupy this huge depression. Eldredge describes it as a fan-shaped system of waterways, grasslands and riverine forest, some 170 kilometers (106 miles) wide and 140 kilometers (87 miles) long—a gigantic oasis where the water remains fresh and life plentiful. The diversity of the vertebrates that coexist here is very great and that of invertebrates even more so.

Dominant in the region's productivity are mound-building termites, *Macrotermes michaelseni*. A mature mound is often three meters high—a dominant feature of the landscape. It contains some five million individual termites, which grow and harvest specialized fungi that occur nowhere else, nourish plants that are abundant only in the vicinity of the mound (some of which form the principal food of the termites) and feed dozens of species of predators large and small. In various stages of maturation and abandonment, the mounds provide homes for many of these animals. In the mounds, an elaborate system of vents maintains the temperature at 25 to 30 degrees Celsius (77 to 86 degrees Fahrenheit) and the humidity between 88 and 96 percent: neither the termites nor the fungi can survive outside these ranges.

Building on this vision of an early and relatively unspoiled paradise, Eldredge shows how humans, with their cattle and fences, are encroaching on the Okavango Delta today. The cattle are barred from the tsetse fly–infested swamps, but the fences halt the native antelopes and other grazing mammals in their normal migratory tracks, and they die in droves along the extensive fence lines. The development of a more comprehensive system, involving both humans and nature, has been undertaken by the Campfire Movement in East Africa and practiced well and beneficially by the Kenya Wildlife Service under the direction of Jonah Western. These approaches point the way to greater success in preserving what Eldredge calls "a fantastic remnant of our own ancestral climes."

Beyond the delta, similar ecosystems have not fared well at the hands of our forebears and are not faring well now. With the cooling of the waters off Africa for the past several million years, the continent itself began to cool and dry out. Savanna vegetation became widespread and largely replaced the rain forests and other dense forests that formerly occupied vast areas. Our own ancestors adopted their distinctive ways of life in these newly formed climes and, Eldredge argues, ultimately came to live in a degree of harmony with the "big hairies" (the large African mammals) that formed such an important feature of their lives. As people spread throughout the world, however, they came into contact with many other big hairies and quickly decimated the populations of these creatures, a ready source of food and sometimes a danger as well.

"As I write these words, we are selling ourselves gasoline at the lowest prices since 1920."

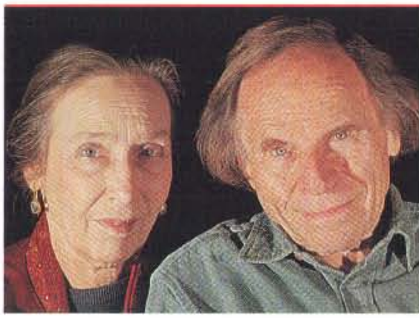
At the time *Homo sapiens* developed agriculture in a number of widely scattered centers, some 10,000 years ago, there were very probably fewer than five million of us throughout the world. Now we number nearly six billion. We consume, waste or divert about half the total primary net photosynthetic productivity on land; we use about half the total supplies of freshwater and affect directly some two thirds of the planet's surface. Over the past 50 years, while our total population has grown by 3.5 billion people, we have lost a quarter of our topsoil and a fifth of our agricultural lands, changed the characteristics of the atmosphere in important ways and cut down a major part of the forests. It is no wonder that we are driving, and will drive, such a high proportion of the other organisms that live here with us to extinction over the course of the next century, thus limiting our own material and spiritual prospects substantially.

In his concluding chapter, Eldredge, having shown the ways in which humans are destroying the very fabric of life on Earth, offers a number of practical suggestions about how we can help ensure its survival in as rich and diverse a form as possible. The first step, he argues persuasively, is to acknowledge the problem. Once we have done so, the attainment of a stable human population at whatever level we choose—thus setting the standard of living that we collectively find acceptable—is one key element in making the transition to a sustainable world. This transition constitutes the most critical task we face in the years to come. We must reformulate our economic principles, employ our existing expertise in conservation and strike a balance between what we perceive as human needs and the continued existence of healthy ecosystems. Perhaps most difficult of all, it means we must organize the political will to confront this problem for the sake of our children and grandchildren—and for the harmonious and productive continuation of life on Earth.

The U.S., wealthiest country the world has ever known, is home to some 4.5 percent of the human population, yet we control about a quarter of the world's economy and cause some 25 to 30 percent of its pollution and ecological damage. This simple relation ought to make us aware that we depend on global stability and on the peace and prosperity of nations everywhere—but we are not. As I write these words, we are selling ourselves gasoline at the lowest prices since 1920. We are deeply in arrears to the United Nations, demanding that weaker and poorer nations take the lead in controlling global climate change. And we have joined a handful of small and politically impotent states in refusing to ratify the Convention on Biological Diversity, already signed by nearly 175 nations and the principal legal instrument for protecting the biodiversity on which we depend on a global basis.

These alarming facts point clearly to the unreality of our worldview and the dangers it poses for us and all other inhabitants of this planet. One can only fervently hope that both this book and the magnificent AMNH exhibit will inspire us, as so many have been inspired in that institution before, to take action on an individual and collective basis to deal with the world as it is rather than to follow blindly the imperfect model that now so dangerously guides our actions.

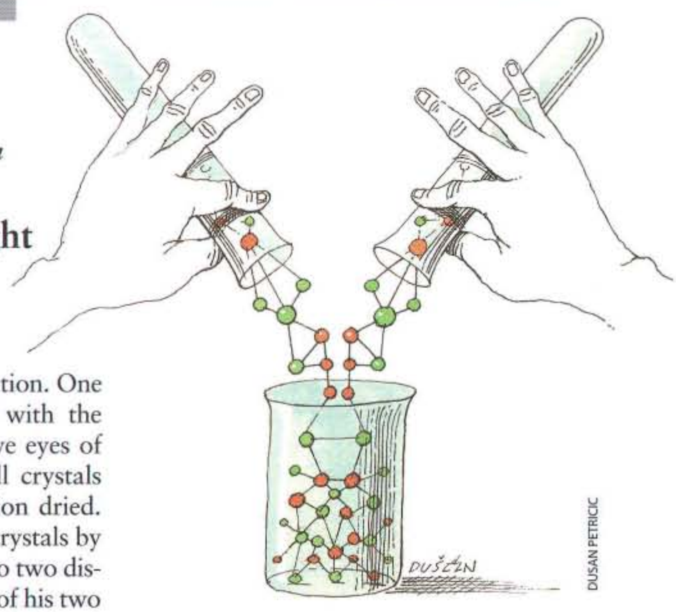
PETER H. RAVEN is director of the Missouri Botanical Garden in St. Louis.



WONDERS

by Philip and Phylis Morrison

The Left and the Right



DUSAN PETRICIC

In 1848 the king of France abdicated his constitutional post—not that he was the only king in Europe to be replaced by republics in those times. Yet the older regime came back; even emperors remained through World War I. That same spring of 1848 an eloquent and enduring pamphlet appeared, the Communist Manifesto of Karl Marx and Friedrich Engels. Its radical appeal would sound throughout the entire 20th century; its critique still defines political Right from Left, even though the demise it prophesied seems farther off than ever after 150 years. The metaphor of right and left is powerful among us still, but at the molecular level, that literal distinction is a rigorous and subtle property of three-dimensional space. And it was in 1848 that its consequences for matter first became clear.

No familiar molecular formula that accounts for the kinds and numbers of atoms present, such as H_2O, is enough to fix its shape fully. All atom-to-atom links must be described geometrically. In a plane that is easy, but in real 3-D it is trickier. A right glove is so similar to the left glove of its pair that a basic description often fits both, although we know that the two gloves cannot equally fit either hand. Each glove matches exactly the image of its partner in an ideal mirror. Mirror-image molecular pairs possess what we call handedness; the chemists term them enantiomers, from the Greek for "opposites."

From grapes and from old wine casks, chemists had by the 1840s separated a number of compounds in crystalline purity. Two similar acids held a mystery. The salts of the two acids—tartaric and racemic—were alike in their chemical properties, yet a solution of tartaric acid's salt rotated the plane of polarized light rightward as a beam passed through, whereas its chemical match, the corresponding salt of racemic acid, was optically inactive, causing no rotation. One postdoctoral chemist gifted with the dexterous hands and attentive eyes of an artist examined the small crystals formed as the inactive solution dried. Then he patiently sorted the crystals by minor differences in shape into two distinct crystal forms, until each of his two piles held enough to be dissolved and measured for rotation. His hand-sorting of the grains was imperfect, yet the rotations he found were opposite in sign and quite similar in amount.

It was this young postdoc, Louis Pasteur himself, who disclosed in May 1848 the strangely dual world of biomolecules. The inactive racemic salt was a mix at the molecular level of two forms: one the expected right-handed salt, the other its lefty antipode. Their rotations canceled in the inactive sample. Pasteur had seen right gloves, left gloves and their 50–50 mix. A recent historian's close study of Pasteur's laboratory notebooks has revealed that his celebrated separation was not at all part of his original experiment design but was more likely a follow-up of visual hints from his own observations.

A certain molecule induces the odor of caraway; in its mirrored form, that of spearmint.

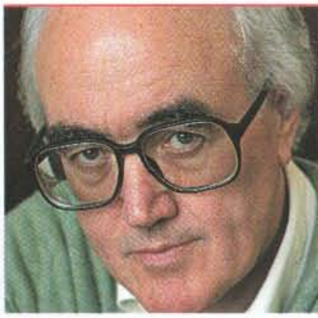
We ourselves own two quartz mineral crystals, each about the size of a quarter-pound of butter; one rotates the plane-polarized light clockwise, the other the other way. Careful inspection of telltale facets and close comparison are convincing: the two big crystals are distinct mirror pairs. But we are dazzled by Pasteur's hundreds of choices by eye and hand at millimeter scale, at once so bold and so tedious.

Early in 1998 we heard a witty and compelling lecture by the chemist K. Barry Sharpless of the Scripps Research Institute. One astonishing accompanying snapshot showed a robust Scripps postdoc holding in each arm a 10-pound bottle of white tartaric crystals. He had prepared the two distinct forms in bulk in a day's work, using catalytic chemical mixtures. The contrast between the postdocs of 1848 and of 1998 at work on the same task offers a thrilling view of the power of today's organic synthesis.

The Sharpless school has found its way over 20 years to a variety of practical catalytic syntheses of handed molecules, usually starting with straight chains of a dozen or so carbon atoms and a couple of side groups, to add a number of wanted groups with the chosen handedness. The necessary information of which molecular glove to sort out is supplied by choosing a natural one-handed derivative of quinine. The catalyst is a metal ion, often highly charged and multivalent. Osmium is the best, a fierce oxidizer for organics. Each expensive osmium-bearing molecule draws around it the handed starter and a few other adherents, forming and unforming that molecular jig tens of thousands of times.

The enzymes of life, all larger protein catalysts, are able to recycle millions of times. They deliver a product scrupulously one-handed to seven decimal places; lab syntheses admit the wrong hand a

Continued on page 83



CONNECTIONS
by James Burke

Tick Tock

Sacre bleu,
Newton had been right!

Walking through Parliament Square in London the other day, I suddenly remembered that when I was a very small child during World War II, the sound of Big Ben ringing the hour on the radio was a comforting indication that it hadn't been hit by incoming German V-1 missiles.

In 1859 the new Big Ben clock became the miracle of harrumph high-tech precision (as had been promised by Prime Minister George Canning) when the gravity escapement system created by E. Beckett Denison was installed. Denison's trick for isolating the movement of the pendulum from that of the gears worked in such a way that no matter how much dirt or ice accumulated on the four sets of clock hands, Big Ben would keep time accurate to the second. (Thank you, Frederickton, New Brunswick, on whose often icy cathedral clock Denison had done a wet run.)

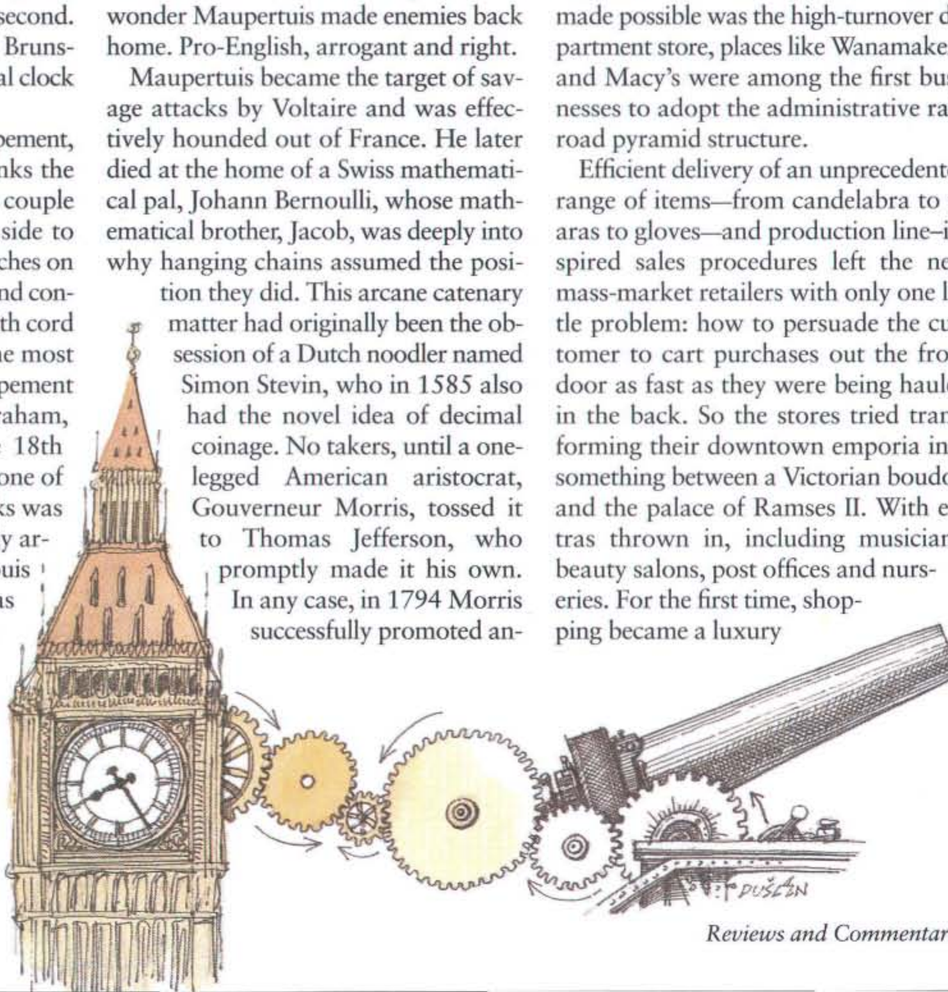
Now, your basic deadbeat escapement, on which Denison improved, links the movement of the pendulum to a couple of blades. As these move from side to side and turn about, each one catches on the clock drive-wheel gear teeth and controls its turn on a shaft wound with cord attached to a weight. Perhaps the most famous version of deadbeat escapement was the brainchild of George Graham, the greatest clockmaker of the 18th century. In the summer of 1736 one of his astronomically accurate clocks was taken to Sweden by a particularly arrogant French scientist, Pierre-Louis Moreau de Maupertuis, who was intent on proving the English right about the earth's shape.

That geodesy should have been at issue was not just another example of the Gallic chip on the shoulder that had been there ever since Newton. It was also the small matter of multiple shipwreck. Get the shape of the earth wrong, and you also get wrong the measure of a degree across its surface. Which means that ships hit rocks that shouldn't be there, because the ships aren't where their navigators think they are. Entire fleets were blundering about in this fatal way and far too often taking to the bottom large quantities of gold and guns. Maupertuis offered a solution: using a one-degree difference in the position of a star (hence the need for precise timing), he calculated exactly how far across the planetary surface represented one degree of latitude. This took him into several months of mosquito hell in the middle of Swedish nowhere. And at the end of it, *sacre bleu,* Newton had been right! The earth was an oblate sphere. No wonder Maupertuis made enemies back home. Pro-English, arrogant and right.

Maupertuis became the target of savage attacks by Voltaire and was effectively hounded out of France. He later died at the home of a Swiss mathematical pal, Johann Bernoulli, whose mathematical brother, Jacob, was deeply into why hanging chains assumed the position they did. This arcane catenary matter had originally been the obsession of a Dutch noodler named Simon Stevin, who in 1585 also had the novel idea of decimal coinage. No takers, until a one-legged American aristocrat, Gouverneur Morris, tossed it to Thomas Jefferson, who promptly made it his own. In any case, in 1794 Morris successfully promoted another scheme (this also snitched, by another politician, DeWitt Clinton): build a canal from Lake Erie to the Hudson. A project famed in song and story until somebody built a railroad across New York State in 1851, which eventually doomed the canal.

By 1855 the New York & Erie Railroad had 4,000 employees and logistical problems so complex that its general superintendent, Daniel McCallum, was obliged to invent modern business administration to deal with them: autonomous heads of division; professional middle management; daily, weekly and monthly reports—all that stuff. And because one of the things regular and high-volume railroad deliveries of goods made possible was the high-turnover department store, places like Wanamaker's and Macy's were among the first businesses to adopt the administrative railroad pyramid structure.

Efficient delivery of an unprecedented range of items—from candelabra to tiaras to gloves—and production line–inspired sales procedures left the new mass-market retailers with only one little problem: how to persuade the customer to cart purchases out the front door as fast as they were being hauled in the back. So the stores tried transforming their downtown emporia into something between a Victorian boudoir and the palace of Ramses II. With extras thrown in, including musicians, beauty salons, post offices and nurseries. For the first time, shopping became a luxury



DUSAN PETRICIC

experience. Once, that is, management had convinced its lady clients (the men wouldn't) to turn up and shop till they dropped.

Wanamaker's gave this crucial public-relations exercise to N. W. Ayer, the first fully fledged advertising agency, and near the end of the 19th century a new breed of thinkers known as psychologists were being recruited to delve deeper into the consumer's emotional motivations to see if further profitable manipulation might be in the cards. Harvard professor of psychology Walter B. Cannon (the only psychologist, I believe, to have had a mountain named after him) advanced matters even more. With the use of the amazing new x-rays and a barium meal (which he invented), Cannon was able to study the peristaltic waves that accompanied digestion and hunger, and he saw that these stopped suddenly if the subject was emotionally affected. After years of experiment, in 1932 Cannon's masterwork, *The Wisdom of the Body,* introduced the concept of homeostasis: the maintenance of balance in the body's state via chemical feedback mechanisms.

Collaborating with Cannon was Mexican neurophysiologist Arturo Rosenbleuth, who in the early 1940s began to have conversations about feedback with a mathematical prodigy, the irascible Norbert Wiener of M.I.T. He was interested in such phenomena as the way feedback works to ensure that when you pick up a glass of water and drink from it, your mouth and glass manage to meet. Wiener was concerned about hits and misses because he was working on the kind of equations that would facilitate antiaircraft artillery in potting their targets more frequently than once every 2,500 shells (the average score on the south coast of England at the time—and no way to win a war).

Wiener's system, called the M-9 Predictor, used feedback from radar data to help forecast where the enemy would be up in the sky when the next shell arrived there. On that basis, his invention directed how the gun should be pointed. The gizmo worked so well that in the last week of large-scale German V-1 attacks, of 104 missiles launched against London only four made it through. Which is why, thanks to Wiener, Big Ben chimed its way, undamaged, right through World War II.

High time I was out of here. SA

Wonders, continued from page 81

few percent of the time. Biological strictness is based on a long-evolved molecular fit for the desired product, a precision key and lock. But the price of life's high precision is a strictly limited start and end. Synthetic catalysis exploits a high-energy intervention instead of the many controlled low-energy steps of protein-led biochemistry, so it can be both selective and versatile.

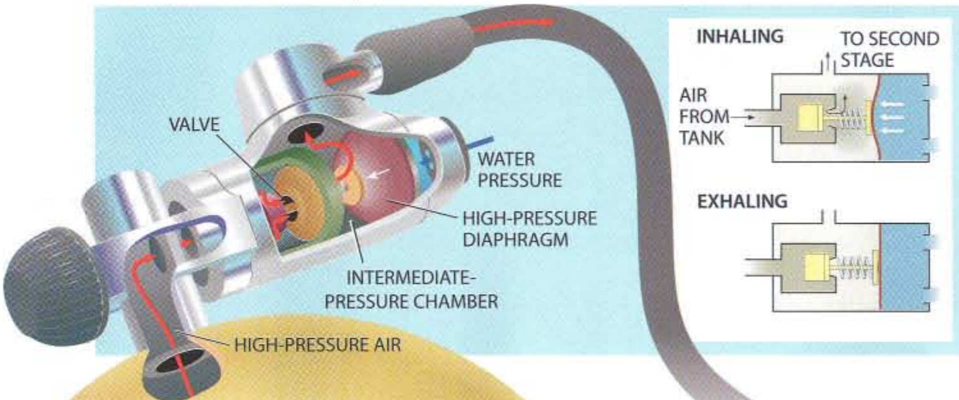
For nonchemists, the insight is that a chemical reaction is not at all only what the brief equation of the reaction shows. Those two balanced sides are merely the endpoints, as though on a transcontinental road trip you saw only Cambridge, Mass., and San Diego, Calif., and nothing of heavy traffic and steep grades between. If we could watch, we would see the heavy-ion catalyst draw its companions, bind and unbind in a sequence of steps, and recycle. Material enough is provided to supply the groups added in the synthesis. A catalytic mix with its auxiliaries includes understandably half a dozen compounds, rationally chosen from a complex lore of reactions and tested by many trials. These are then added in sequence to a big pot of cooled alcohol solvent that is easy to evaporate: no grain-by-grain sorting here!

Almost all hormones, pheromones and other neurological drugs—indeed most specific pharmaceutical or agrobiological compounds—are functional in single-handed forms. But synthetics come as a mix, unless they have been made through the use of enzymes chosen from life or by one of the new asymmetric syntheses. The mirrored molecule can surprise, too: a certain molecule induces the odor of caraway; in its mirrored form, that of spearmint. The wrong hand of Ritalin, a molecular species used or overused to control the hyperactivity of children, seems to be completely inert. It is said that the tragic congenital injuries from thalidomide were in fact a side effect of the worse than useless wrong form. Thus, at best, what is sold at high price is 50 percent inactive.

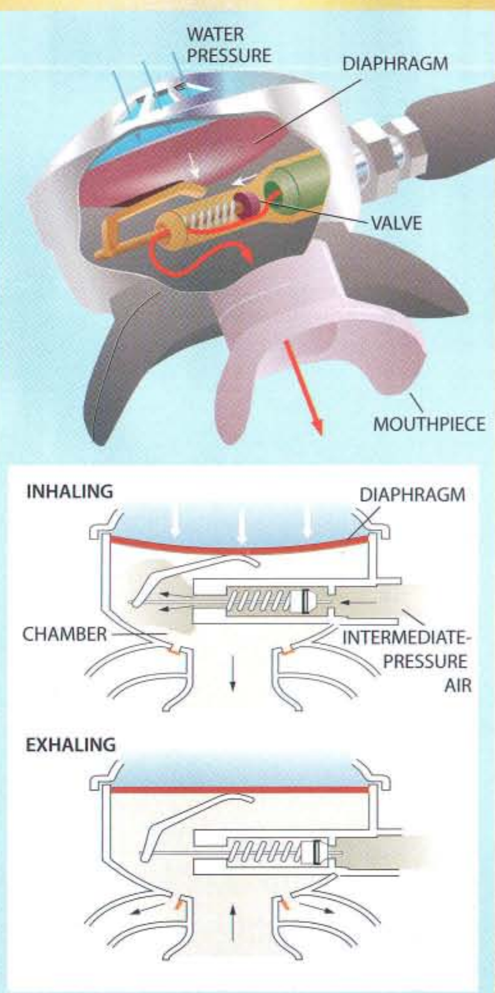
But the synthetics industry is taking swift steps to deliver life products only in the hand desired. Within a decade most handed products will be in your drugstore in one form only; the Food and Drug Administration will be watching its mirrors carefully. SA

WORKING KNOWLEDGE

SCUBA REGULATORS



REGULATOR FIRST STAGE maintains an intermediate pressure of 90 to 140 pounds per square inch in the hose leading to the second stage. There are several different designs; perhaps the most successful is the balanced diaphragm. Whenever the diver takes a breath, the pressure falls in the hose and intermediate-pressure chamber, causing the high-pressure diaphragm to flex inward. This flexing opens a valve and allows high-pressure air to flow from the tank into the chamber and hose. The surrounding (ambient) water pressure against the diaphragm also ensures that the intermediate pressure in the chamber and hose stays at a certain level above the ambient pressure.



REGULATOR SECOND STAGE reduces the intermediate pressure in the hose to the ambient pressure at whatever depth it happens to be, allowing the diver to breathe comfortably at that depth. Like the first stage, it relies on a diaphragm. Breathing in through the mouthpiece causes the diaphragm to flex inward, pushing a lever that opens a valve to admit intermediate-pressure air from the hose. Water pressing against the outer side of the diaphragm matches air pressure in the chamber to the surrounding water pressure.

ILLUSTRATIONS BY KARL GUDE; SCHEMATICS BY BRYAN CHRISTIE

by William A. Bowden
Chief Operating Officer, Dacor Corporation

The demand regulator, which connects a high-pressure scuba tank to a breathing mouthpiece, opened the vast world beneath the waves to hundreds of millions of people. It is not much to look at, but this remarkable device enables anyone in reasonably good health to see what people could only imagine or speculate about for thousands of years. In so doing, it transformed perceptions of the undersea domain from that of a mysterious place having dark and often fearsome connotations to a more familiar realm known mainly for its beauty and wonder.

A regulator enables a diver to breathe comfortably underwater by regulating the tank's high-pressure breathing gas, matching the gas pressure to the ambient pressure at whatever depth the diver happens to be. This regulation takes place in two stages. The first stage reduces the tank pressure, which is typically about 3,000 pounds per square inch (psi) in a full tank, to an intermediate pressure between 90 and 140 psi. The second stage then reduces the intermediate pressure to the ambient pressure—44 psi at a depth of 66 feet in seawater, for example, or 59 psi at 100 feet.

Regulators first became available to the general public in the U.S. in 1953, when Dacor introduced the C1 and U.S. Divers came out with its model of the Aqua-Lung. Regulators metamorphosed significantly in the mid-1960s, when double-hose designs were replaced by single-hose types in which the exhaled gas was discharged near the mouthpiece. Another period of rapid improvement occurred in the late 1970s, when natural rubber and brass components were replaced with more durable and impervious synthetic materials (such as silicone and polyurethane) and stainless steel. Also, there were ergonomic improvements to mouthpieces and exhaust manifolds.